5-1998

Criterion-Related Validity of the Borg Ratings of Perceived Exertion (RPE) Scale : A Meta-Analysis

Michael J. Chen
Utah State University

CRITERION-RELATED VALIDITY OF THE BORG RATINGS OF PERCEIVED EXERTION (RPE) SCALE: A META-ANALYSIS

by

Michael J. Chen

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Psychology

Approved:

UTAH STATE UNIVERSITY
Logan, Utah

1998

ABSTRACT

Criterion-Related Validity of the Borg Ratings

of Perceived Exertion (RPE) Scale:

A Meta-Analysis

by

Michael J. Chen, Doctor of Philosophy

Utah State University, 1998

Major Professor: Dr. Xitao Fan
Department: Psychology

The Borg Ratings of Perceived Exertion (RPE) Scale has proven to be a highly popular instrument in measuring the subjective responses of individuals to a given work or exercise task. Historically, the instrument was designed to correlate highly with the heart rates in young-to-middle-aged men performing various tasks. The body of literature, however, has revealed inconsistencies in the extent of just how strong the relationship is between ratings of perceived exertion and various physiological criterion variables, most notably, heart rate. In addition, most studies have invoked the question of whether the criterion-related validity coefficients derived from the relationship between ratings of perceived exertion and a specified physiological criterion variable are just as valid as those for which the Borg RPE Scale was originally performed. A meta-analysis, therefore, was undertaken to determine the magnitude of the relationship between ratings of perceived exertion scores and each of three commonly used physiological measures or criterion variables: heart rate, blood lactate, and oxygen uptake.

Results show that by using Tests of Homogeneity for each physiological criterion variable, the observed sample size-weighted validity coefficients are heterogeneous. The median of the mean sample size-weighted validity coefficients is .574 for heart rate, .561 for blood lactate, and .480 for oxygen uptake. Each study in the meta-analysis was grouped by the study characteristics of subject gender, fitness level, RPE Scale, exercise type, exercise protocol, and study quality. For heart rate, the highest validity coefficients are those in which the subjects are highly fit, the exercise type is fairly unusual, such as swimming, and the subjects are required to maximally exert themselves. For blood lactate, the highest validity coefficients are for females, healthy-inactive subjects, the 15-point RPE Scale, treadmill use, and swimming. For oxygen uptake, the highest validity coefficients between ratings of perceived exertion and oxygen uptake are for swimming.

In a meta-analysis of study effects, when the validity coefficients are analyzed by study, the resultant mean validity coefficients are only somewhat higher (ratings of perceived exertion and heart rate, .657; ratings of perceived exertion and blood lactate, .642; ratings of perceived exertion and oxygen uptake, .609) than those obtained using sample size-weighted validity coefficients.

Finally, corrections for bias generally resulted in increased validity coefficients and decreased variances.

(163 pages)

DEDICATION

This work is dedicated to my wife, Sondra Telford Moe, whose yin to my yang, and whose strength of character, munificence, love, and support have been my guiding light.

ACKNOWLEDGMENTS

For his invaluable guidance, advice, patience, and counsel, a huge debt of thanks and gratitude go to my major professor, Dr. Xitao Fan. I also owe a huge debt of thanks and gratitude to Dr. Blaine Worthen for showing his faith in me by appointing me to the position of assistant editor of Evaluation Practice (now, The American Journal of Evaluation). I would also like to thank my committee members, Drs. Keith Checketts, Donald Sisson, Lani Van Dusen, and Blaine Worthen, for their support and assistance.

Michael J. Chen

CONTENTS

Page

ABSTRACT . . . iii

DEDICATION . . . v

ACKNOWLEDGMENTS . . . vi

LIST OF TABLES . . . ix

LIST OF FIGURES . . . xii

CHAPTER

I. GENERAL INTRODUCTION AND PROBLEM STATEMENT . . . 1

II. REVIEW OF THE LITERATURE . . . 4

Perceived Exertion and Its Measurement . . . 4
Meta-Analysis . . . 16

III. PURPOSE AND OBJECTIVES . . . 33

IV. METHOD . . . 35

Study Features . . . 36
Statistics . . . 40
Study Effects Meta-Analysis . . . 48

V. RESULTS . . . 49

Test of Homogeneity . . . 49
Mean Validity Coefficients . . . 51
Correction for Artifacts . . . 75
Study Effects Meta-Analysis . . . 78

VI. DISCUSSION AND CONCLUSIONS . . . 79
Test of Homogeneity . . . 79
Mean Validity Coefficients . . . 90

Correction for Artifacts 97
Study Effects Meta-Analysis 99
General Conclusions and a Recommendation for Future Directions 101

REFERENCES 105

APPENDIX 120

CURRICULUM VITAE 145

LIST OF TABLES

Table		Page
1	Features of Studies	37
2	Study Artifacts That Cause the Study Validity Coefficient to Differ from the True (Population) Correlation and the Corrections Associated with These Artifacts	46
3	Homogeneity Test for Correlations Between Ratings of Perceived Exertion and the Three Criterion Variables	50
4	Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Heart Rate by Gender	52
5	Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Heart Rate by Fitness Level	53
6	Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Heart Rate by RPE Scale	54
7	Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Heart Rate by Exercise Type	55
8	Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Heart Rate by Exercise Protocol	56
9	Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Heart Rate by Study Quality	57
10	Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Blood Lactate by Gender	59
11	Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Blood Lactate by Fitness Level	60
12	Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Blood Lactate by RPE Scale	61
13	Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Blood Lactate by Exercise Type	62

14 Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Blood Lactate by Exercise Protocol . . . 63

15 Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Blood Lactate by Study Quality . . . 64

16 Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Oxygen Uptake by Gender . . . 66

17 Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Oxygen Uptake by Fitness Level . . . 67

18 Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Oxygen Uptake by RPE Scale . . . 68

19 Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Oxygen Uptake by Exercise Type . . . 69

20 Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Oxygen Uptake by Exercise Protocol . . . 70

21 Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Oxygen Uptake by Study Quality . . . 71

22 Statistically Significant Two-Way Interactions for Heart Rate and Oxygen Uptake . . . 73

23 Artifact-Corrected Validity Coefficients of Each Study Feature for the Three Criterion Variables . . . 76

24 Mean Validity Coefficients for the Study Effects Meta-Analysis . . . 78

A.1 Coding of Characteristics and Validity Coefficients for Ratings of Perceived Exertion and Heart Rate . . . 121

A.2 Coding of Characteristics and Validity Coefficients for Ratings of Perceived Exertion and Blood Lactate . . . 131

A.3 Coding of Characteristics and Validity Coefficients for Ratings of Perceived Exertion and Oxygen Uptake . . . 134

A.4 Descriptive Statistics for Weighted Correlations for Heart Rate . . . 138

A.5 Descriptive Statistics for Weighted Correlations for Blood Lactate 139

A.6 Descriptive Statistics for Weighted Correlations for Oxygen Uptake 140

A.7 Descriptive Statistics for Statistically Significant Two-Way Interactions by Study Characteristics for Heart Rate 141

A.8 Descriptive Statistics for Statistically Significant Two-Way Interactions by Study Characteristics for Oxygen Uptake 142

A.9 Mean Validity Coefficients by Criterion Variable and Study (Overall) 143

LIST OF FIGURES

Figure		Page
1	The Borg 15-point Ratings of Perceived Exertion (RPE) Scale	7
2	Stem-and-leaf plot and boxplot showing the distribution of validity coefficients for heart rate	83
3	Stem-and-leaf plot and boxplot showing the absence of outliers for validity coefficients for blood lactate	83
4	Stem-and-leaf plot and boxplot showing the distribution of validity coefficients for oxygen uptake	84

CHAPTER I

GENERAL INTRODUCTION AND PROBLEM STATEMENT

Because regular exercise is a potent means for maintaining a healthy lifestyle (Burlew, Jones, & Emerson, 1991), particularly as the body grows older, it is essential that individuals are able to monitor the intensity of such exercise. Generally, validation of instruments designed to measure exercise or work intensity is essential for all age groups and occupations.

Perhaps because of its ease of administration, the Borg Ratings of Perceived Exertion (RPE) Scale is probably the most prevalent psychological measure of perceived exertion or exercise intensity in use today. This psychological measure is a method to determine the intensity of effort, stress, or discomfort that is felt during exercise and is correlated with pain threshold or tolerance (Morgan, 1973). Research related to the Borg RPE Scale has focused on how it is related (correlated) with a wide variety of physiological measures (e.g., heart rate, ventilatory drive, blood lactate levels, creatine levels) and psychological measures (e.g., health questionnaires, surveys, diaries) and dispositions (e.g., personality types). Unfortunately, this research has yielded apparently inconsistent results. For example, although the Borg RPE Scale was originally designed to be related to heart rate, not all studies conducted since then have found the same strong correlation with heart rate as the original study did. Likewise, there is a great deal of inconsistency about the relationship between ratings of perceived exertion and blood lactate levels.

Such lack of consistency undermines the validity of the use of the Borg RPE

Scale in general. The inconsistent evidence relating one criterion variable (e.g., heart rate) with RPEs may reflect the imperfections of individual studies themselves. Indeed, "it is perhaps characteristic of this field that workers using different techniques to answer the same question produce conflicting results" (Cafarelli, 1982, p. 386). Since no study is ever conducted perfectly, the validity coefficient from an individual study cannot be used to directly estimate the true criterion-related validity of RPE. Instead, the validity generalization based on a body of literature requires some statistical treatment. And therein lies the goal of meta-analysis for such validity generalization: to describe quantitatively the distribution of criterion-related validity coefficients of ratings of perceived exertion.

Although the plethora of research related to the Borg RPE Scale has been extensively reviewed during the past two decades (see Noble & Robertson, 1996), none of these reviews has addressed the use of the Borg RPE Scale as it relates to physiological measures in a quantitative fashion. Rather, all are narrative summaries and/or syntheses. There is a need, therefore, to quantitatively synthesize the relevant literature pertaining to inconsistencies about the relationship of ratings of perceived exertion with criterion variables such as heart rate, blood lactate levels, and oxygen uptake. The ability of meta-analysis to integrate findings across studies, organize the available validity coefficients into a distribution of coefficients for a particular relationship (such as that between heart rate and ratings of perceived exertion), as well as its power to reduce sampling error, while at the same time, correct for artifacts, makes it an ideal technique to use for quantifying the empirical relationship between ratings of perceived exertion and criterion variables. Such quantification will underscore what Hunter and Schmidt (1990) call "validity generalization," which is defined as the extent to which the measurement of a construct (in this case, perceived exertion or exercise intensity) is externally valid beyond the samples from which the coefficients were originally derived.

The purpose of this study, therefore, is to conduct a meta-analysis of the extant literature that has empirically examined the criterion-related validity evidence of perceived exertion with such criterion variables as heart rate, blood lactate, and oxygen uptake. These three criterion variables related to ratings of perceived exertion dominate the perceived exertion literature and are, therefore, not only the most amenable to a meta-analysis, but are also in the highest need for it.

CHAPTER II

REVIEW OF THE LITERATURE

Perceived Exertion and Its Measurement

It is now widely accepted that a regular exercise regimen promotes a healthy lifestyle (Burlew et al., 1991; King & Senn, 1996; McArdle, Katch, & Katch, 1996; Moriya & Fukuchi, 1990), reduces the risk of disease (Blair et al., 1989; Holloszy, 1990; King & Senn, 1996) and injury (Ibusuki, Kondo, Soya, & Yagi, 1990; Kobayashi, Hirano, & Fukunaga, 1990), enhances cognitive functions and emotional dispositions (Dustman et al., 1990; Emery & Blumenthal, 1991; McAuley & Rudolph, 1995; Plante & Rodin, 1990), and may even increase longevity (Rakowski & Mor, 1992). And as the body ages, and as the importance of regular exercise increases, it becomes likewise increasingly more crucial for an individual to accurately monitor his/her exercise progress. However, the measurement of the intensity of effort, stress, or discomfort that is felt during exercise and that is correlated with pain threshold or tolerance (Morgan, 1973) is often accomplished indirectly. Perceived exertion, therefore, has become the most frequently used proxy for measuring exercise intensity.

There is an increasing trend for fitness and health professionals in telling people to monitor their perceptions of effort to determine their exercise intensity, in addition to the more traditional method of monitoring heart rate (American College of Sports Medicine, 1991). With the exception of populations requiring more precise knowledge of heart rate, such as postinfarct patients, the use of perceived effort has become a prevalent tool for monitoring exercise intensity (see Noble & Robertson, 1996). One reason for using perceived effort as an intensity guide is that attaining a specified heart rate zone may not always be necessary or desirable, especially if the primary goal is to promote regimen adherence. Unlike heart rate monitoring, teaching individuals to "listen" to their bodies allows for periodic fluctuations in physiological and psychological responses that occur during exercise. During the adoption phase of an exercise regimen, an individual who is simultaneously determined to reach a specified heart rate and having a "bad exercise day" is likely to perceive the bout to require much more effort and thus experience a considerable degree of negative affect.

Prevalence and Application of Perceived Exertion

For many years, applied physiologists and psychologists have realized that the human decision to continue or cease hard physical work, as well as the intensity at which a person chooses to work, is governed in large part by the person's subjective feelings. So important, in fact, was the concept of perceived exertion, that two symposia devoted exclusively to research and clinical applications of perceived exertion were organized about a decade apart (1972 and 1981; Noble, 1982). Perceived exertion has been applied to a wide variety of fields such as environmental factors, exercise and prescription, occupational factors, ergogenic aids, pulmonary functions, growth, aging, and gender, and many others, and has also been the focus of 13 major review articles since 1970 (Noble & Robertson, 1996). Not one of these reviews, however, is a quantitative synthesis (i.e., meta-analysis). More recently, a comprehensive monograph was published (Noble & Robertson, 1996) about the historical development of perceived exertion, the development, administration, and experimental use of the Borg RPE Scale (Figure 1), and psychological, physiological, and perceptual mediators in the effort sense.

The review of the literature here will not recount what has been covered by Noble and Robertson, but relevant literature since then will be reviewed briefly. More importantly, the review here will focus on the literature pertaining to the validity of the use of the Borg RPE Scale.

Measurement of Perceived Exertion

In sports psychology and exercise science, measurement follows basically all the same principles, protocols, and procedures, and faces many of the same problems (Safrit, 1989) often encountered in educational and psychological measurement settings, but with the additional physiological/physical variables being taken into account. Central to the concept of measurement is the issue of validity, which is often more difficult to establish, methodologically, than reliability (Sim & Arnell, 1993). One of the most intensely studied clinical applications of perceived exertion is that of physical rehabilitation and exercise prescription (Dishman, 1994; Dunbar & Bursztyn, 1996; Dunbar, Glickman-Weiss, Edwards, Conley, & Quiroz, 1996; Dunbar, Kalinski, & Robertson, 1996; Noble & Robertson, 1996; Parfitt & Eston, 1995). For these purposes, valid use of the Borg RPE Scale has become a critical issue.

6	
7	Very, very light
8	
9	Very light
10	
11	Fairly light
12	
13	Somewhat hard
14	
15	Hard
16	
17	Very hard
18	
19	Very, very hard
20	

Figure 1. The Borg 15-point Ratings of Perceived Exertion (RPE) Scale

Because the Borg RPE Scale was designed as a proxy indicator of exercise intensity, the most relevant validity evidence for perceived exertion is criterion-related validity evidence, which describes the empirical relationship between ratings of perceived exertion with some physiological variables more directly related to exercise intensity. The three most common criterion variables used in research related to perceived exertion are heart rate, blood lactate, and oxygen uptake.

The Borg RPE Scale as an Instrument to Measure Perceived Exertion

Historically, the concept of perceived exertion was originated by Borg in the early 1960s and rapidly gained momentum during the next decade (Noble & Robertson, 1996). The Borg 15-point RPE Scale (Figure 1) was developed so that perceptual ratings increase linearly with power output and heart rate (Borg & Linderholm, 1967). In fact, despite several findings in the literature to the contrary (Noble & Robertson, 1996), any physiological variable linearly related to exercise intensity tends to parallel perceptual ratings obtained from the Borg RPE Scale; however, certain physiological variables, especially lactate production, are related to exercise intensity according to nonlinear power functions (Borg, Hassmen, & Lagerstrom, 1987; Borg, Van den Burg, Hassmen, Kaijser, & Tanaka, 1987).

Limitations of the Borg RPE Scale. From a psychometric point of view, the Borg RPE Scale is extremely limited since it is only one scale (or one item, see Figure 1), rather than a (full-length) questionnaire with many items, that, therefore, flies in the face of the current belief that "a longer look gives greater accuracy" (Rogosa & Ghandour, 1991, p. 282). Another potent limitation of the Borg RPE Scale is the presence of psychological and physiological confounding variables (Rejeski, Hardy, & Shaw, 1991). But in exercise (self-) prescription, such psychological confounding variables can be of beneficial use, as in exercise self-monitoring (Evans, Hopkins, & Toney, 1996). Mood or affect (Parfitt & Eston, 1995; Parfitt, Eston, & Connolly, 1996) and exercise history (Parfitt et al., 1996; Parfitt, Markland, & Holmes, 1994) have recently been shown to be a possible confounding variable and interfere with higher ratings of perceived exertion.

External validity. In keeping with the sentiments expressed by Messick (1995) "that validity is an evolving property and validation a continuing process" (p. 741), comparable results of two or more studies using different kinds of subjects and protocols, but leading to the same conclusions, would serve to strengthen the generalizability of those results. For example, if differences in a procedural detail, such as whether the Borg RPE Scale is visible or invisible to subjects (Abadie, 1996), lead to the same results and conclusions, the external validity of the results and their subsequent meaning is strengthened. Several other investigators have recognized differences between their protocols and those of earlier studies focusing on (roughly) the same issue, such as examination of the relationship between perceived exertion and work load (Morgan, 1973; Skinner, Hustler, Bergsteinova, & Buskirk, 1973a), circadian rhythms (Trine & Morgan, 1995), exercise self-efficacy (Rudolph & McAuley, 1996), power output (Robertson et al., 1996), or local (peripheral) lactate accumulation (Noble, Borg, Jacobs, Ceci, & Kaiser, 1983; Noble & Robertson, 1996), as well as with a wide variety of other parameters (Noble & Robertson, 1996). Affect (Hardy & Rejeski, 1989; Martin & Anshel, 1995) or personality type (A vs. B) can also act as modifiers of perceived exertion (Hassmen & Koivula, 1996). Indeed, "it is perhaps characteristic of this field that workers using different techniques to answer the same question produce conflicting results" (Cafarelli, 1982, p. 386).

The use of the Borg RPE Scale has been validated using young male and female adults on the bicycle ergometer mostly against heart rate measurements (Dunbar et al., 1992; see Noble & Robertson, 1996; Potteiger & Evans, 1995; Skinner et al., 1973a; Skinner, Hustler, Bergsteinova, & Buskirk, 1973b), percent maximal oxygen uptake (Dunbar & Bursztyn, 1996; Dunbar, Kalinski et al., 1996; Dunbar et al., 1992), blood and muscle lactate levels (Noble & Robertson, 1996), and on walking tasks (see above).

Noble and Robertson (1996) recognized that the majority of evidence linking heart rate with perceptual signals of exertion was derived from correlational data. Borg's (1962) initial efforts to validate the Borg RPE Scale yielded a correlation of 0.85 between heart rate and ratings of perceived exertion to progressively increasing power outputs. Correlation coefficients ranging from $r = 0.42$ to $r = 0.94$ have been found between heart rate and ratings of perceived exertion while lifting weights, pushing a wheelbarrow, riding a cycle ergometer, transporting external weights, treadmill walking, performing one- and two-limb exercises, and being immersed in cold water (Noble & Robertson, and references cited therein). The relationship between heart rate and ratings of perceived exertion is approximately the same in younger and older age groups (Eston & Williams, 1986).

Recent alterations of the Borg RPE Scale--validity of their use and interpretation. The Borg RPE Scale has undergone several modifications, such as the Three-Point Method and the Slope Method. The Three-Point Method entails plotting ratings of perceived exertion as a function of oxygen uptake (VO_2) using three data points. Point 1 is rating 6 and $VO_2 = 3.5$ ml kg^{-1} min^{-1}. Point 2 is rating 13 at which an estimate of VO_2 is determined by asking subjects to select an exercise intensity on the treadmill (or other exercise apparatus) which they perceive to be an rating of 13. The selected treadmill speed is then used to estimate VO_2 using the American College of Sports Medicine metabolic calculation equations (ACSM, 1991). Point 3 is rating = 20 and $VO_{2\,max}$. The ratings of perceived exertion equivalents of 50%, 60%, 70%, and 85% of maximal oxygen uptake ($VO_{2\,max}$) are then determined from the resultant plot. The Three-Point Method allows development of accurate ratings of perceived exertion exercise prescriptions in numerous fitness and clinical settings wherein previous techniques were not applicable; this technique has been validated for healthy individuals (Dunbar, Kalinski et al., 1996), and cardiac and pulmonary patients (Dunbar & Bursztyn, 1996).

The Slope Method entails the use of a set of equations to establish the slope of a line describing the relationship between oxygen consumption and ratings of perceived exertion; subsequently, this line is used to find the estimated VO_2 peak and the ratings of perceived exertion associated with various percentages of the VO_2 peak. The Slope Method, which requires only a submaximal exercise test, minimizes error, increases utility of the procedure, is technically and methodologically simpler than the Three-point Method and, unlike the latter, it does not entail graphing and can be used to estimate $VO_{2\,max}$ (Dunbar & Bursztyn, 1996). Nevertheless, both the Three-point Method and the Slope Method (Dunbar & Bursztyn, 1996) are valid for developing exercise prescriptions based on ratings of perceived exertion.

Three Criterion Variables Used to Measure
Perceived Exertion

Heart rate. Historically, because heart rate measurements were initially used to validate the Borg RPE Scale (Borg, 1973, 1982; Borg & Linderholm, 1967, 1970), they can be useful indices of perceived exertion (Bar-Or, Skinner, Buskirk, & Borg, 1972; Kolkhorst, Mittelstadt, & Dolgener, 1996; Miller, Bell, Collis, & Hoshizaki, 1985; Netz, 1987; Sidney & Shephard, 1977; Travlos & Marisi, 1996; Wong, Cunningham, Rechnitzer, & Howard, 1990; Zeni, Hoffman, & Clifford, 1996) and may even substitute for the Borg RPE Scale in field experiments employing graded exercise tasks (Potteiger & Evans, 1995), depending on the exercise modality used (Carton & Rhodes, 1985; Garcin, Cravic, Vandewalle, & Monod, 1996; Zeni et al., 1996). From a practical point of view, among the three criterion variables reviewed in this review of the literature, heart rate is the easiest to measure. And unlike blood lactate concentration that provides a local or peripheral measure of exertion, heart rate measurements and oxygen uptake provide more global measures of cardiovascular fitness and are, therefore, subject to a variety of potent confounding variables, such as the individual's emotional state, and may not be necessarily measuring (only) perceived exertion.

Blood lactate concentrations. Because changes in blood lactate concentrations are the result of more specific or local (peripheral) influences (Noble & Robertson, 1996), such as a predominance of anaerobic metabolism as a consequence of highly (over) worked skeletal muscle in the presence of low oxygen levels, blood lactate concentrations provide a useful objective index of exercise intensity and perceived exertion (Billat, 1996; Carton & Rhodes, 1985; Dishman, Farquhar, & Cureton, 1994; Hetzler et al., 1991; Noble & Robertson, 1996; Prusaczyk, Cureton, Graham, & Ray, 1992; Seip, Snead, Pierce, Stein, & Weltman, 1991; Steed, Gaesser, & Weltman, 1994; Stoudemire et al., 1996; Zeni et al., 1996), and are not influenced by emotional states (Dishman et al., 1994; Seip et al., 1991) as is the more global condition of cardiovascular fitness, although it is still possible for genetics to have a(n) (minor) influence, such as a nonlethal mutation in, say, lactate dehydrogenase.

Furthermore, from a practical point of view, if the determination of blood lactate levels is not performed routinely in a particular laboratory, it is not usually the method of choice. Since it is invasive, expensive, and technically demanding, it has not gained widespread use.

Oxygen uptake. Unlike blood lactate concentration, which provides a local or peripheral measure of exertion, heart rate measurements and oxygen uptake provide more global measures of cardiovascular fitness and are, therefore, subject to a variety of potent confounding variables, such as the individual's emotional state, and may not be necessarily measuring (only) perceived exertion. Thus, in general, experimental evidence indicates that for most exercise conditions, the relative metabolic rate functions as a mediator for respiratory-metabolic signals of exertion (Noble & Robertson, 1996).

The Relationship Between RPE and the Three Criterion Variables

Heart Rate

There is much correlational evidence (e.g., Noble & Robertson, 1996) suggesting that heart rate may function as a perceptual signal mediator. On the other hand, a substantial amount of experimental evidence shows a general lack of correspondence between heart rate and ratings of perceived exertion when one or the other variable has been experimentally manipulated during dynamic exercise. For example, Pandolf (1983) demonstrated that during submaximal exercise, heart rate increased as a function of increasing environmental temperature, while ratings of perceived exertion was unchanged by heat stress. Other researchers (Ainsworth, McMurray, & Veazey, 1997; Smolander, Korhonen, & Ilmarinen, 1990; Takeshima et al., 1996) found that heart rate measurements cannot always be used to predict exercise intensity whether subjects are trained or untrained. Thus, ever since the initial attempts at validation using heart rate (Borg, 1973, 1982), the literature has indicated several inconsistencies in associating heart rate measurements with perceived exertion (see below; Carton & Rhodes, 1985; Noble & Robertson, 1996). Thus, when the experimental and correlational evidence is examined collectively, it can be concluded that heart rate does not appear to function as a physiological mediator for respiratory-metabolic signals of exertion.

Blood Lactate

Ratings of perceived exertion are known to be highly dependent on whether the perceived effort is derived from the extremities (peripheral factors) and the size of the muscle groups involved (Aminoff, Smolander, Korhonen, & Louhevaara, 1996; Borg et al., 1987; Hoffman, Kassay, Zeni, & Clifford, 1996) or from the torso (central factors) (Cafarelli, 1982; Noble & Robertson, 1996; Robertson, 1982).

The common difficulty of subjects to differentiate between peripheral versus central exertion does much to contribute to the inconsistent results found in the literature (Carton & Rhodes, 1985; Noble & Robertson, 1996). Thus, whether blood lactate and RPE are positively related (Borg, Ljunggren, & Ceci, 1985; Demello, Cureton, Boineau, & Singh, 1987; Gamberale, 1972; Horstman, Morgan, Cymerman, & Stokes, 1979), or not (Allen, Seals, Hurley, Ehsani, & Hagberg, 1985; Robertson, Gillespie, McCarthy, & Rose, 1979; Skrinar, Ingram, & Pandolf, 1983; Stamford & Noble, 1974) is still unknown.

Oxygen Uptake

The functional link between (percent maximal) oxygen uptake and ratings of perceived exertion has important implications for perceptually regulated exercise prescriptions, as well as for laboratory assessment of exercise performance. That is, establishing a link between (percent maximal) oxygen uptake and ratings of perceived exertion is a prerequisite for perceptually prescribed exercise intensity requiring the subject to estimate his/her amount of exercise intensity (Noble & Robertson, 1996). Fortunately, heart rate measurements have also been shown to be highly correlated with percent maximal oxygen uptake (Garcin et al., 1996; Haskell, Yee, Evans, & Irby, 1993), which has been shown to be one of the most valid measures of physical fitness (Brown, Chitwood, Beason, & McLemore, 1996a, 1996b; Liu, Plowman, & Looney, 1992; Murray et al., 1993; Siconolfi, Lasater, Snow, & Carleton, 1985). Oxygen uptake has been shown to be correlated with ratings of perceived exertion (r = 0.76 to 0.97; e.g., Goslin & Rorke, 1986; Toner, Drolet, & Pandolf, 1986), although, again, not all studies agree (DeMello et al., 1987; Pivarnik & Senay, 1986).

Summary

To quote Caferelli (1982) once more, "it is perhaps characteristic of this field that workers using different techniques to answer the same question produce conflicting results" (p. 386). This brief quote underscores the general theme of this literature review. And nowhere is this more evident than in the amount of inconsistent evidence in linking ratings of perceived exertion to blood lactate and, to a lesser extent, to heart rate. Although a large majority of the studies reviewed found a positive relationship between ratings of perceived exertion and heart rate, they did not all agree on the magnitude of the correlation; the same could be said of the correlation between ratings of perceived exertion and oxygen uptake. Such lack of agreement among studies examining the relationships between ratings of perceived exertion and criterion variables undermines the overall validity and reliability of the inferences made regarding the use of the Borg RPE Scale. These inconsistencies, however, can be partially resolved through the use of quantitative meta-analytic syntheses, which will help to determine the extent to which ratings of perceived exertion are related to such criterion variables as heart rate, blood lactate, and oxygen uptake.

Meta-Analysis

Meta-analysis employs a set of statistical methods as tools for the quantitative integration of research findings across studies. Meta-analysis is generally not considered as a statistical technique; instead, it is a quantitative approach for integrating the empirical results of many independent studies. The subject domain of meta-analysis itself, however, is not without its share of controversies, some of which will be discussed momentarily. As will be reviewed below, there are different forms of quantitative integration of research results, and only some of them can be called meta-analysis.

Relevant to the meta-analysis of the validity coefficients reported in the Borg RPE Scale literature, the methods for synthesizing the results of correlational studies that report on the relationship between ratings of perceived exertion and some physiological or psychological measure are relatively straightforward. But even here controversy abounds. One can either (a) average the raw Pearson rs obtained from each study (Wolf, 1986), or (b) transform each r into its associated z statistic using Fisher's r-to-z transformation that is then averaged and then transformed back to r (Wolf, 1986). But even here, which method of integration to choose is the subject of some controversy (Glass, 1977; Wolf, 1986).

A Brief Account on the Origin of Meta-Analysis

Meta-analysis and its competing or associated techniques were born out of a dire need to make some quantitative sense out of the explosive growth the social science and educational research literature experienced during the seventies (Glass, 1977). Before then, a literature review typically covered a few dozen articles, depending on the subject domain in question. Because of its relatively modest scope at the time, in writing a literature review it would be easier for an author to distill and verbally summarize the cumulated studies in a rather coherent fashion. But this is not to say that a narrative review cannot be confusing even if only a small number of studies are reviewed. Today, however, although verbal narratives and descriptions are obviously still possible, narrative reviews become increasingly difficult and subjective because of the ever-increasing body of literature. Indeed, Hunter and Schmidt's (1990) monograph focuses on

> the cumulation of results across studies to establish facts,....the resolution of the basic facts from a set of studies that all bear on the same relationship. For many years, this was not an important issue in the social sciences because the number of studies dealing with a given issue was small. But that time has passed. (p. 13)

The rapid growth in the number of studies, therefore, has served as the impetus for meta-analysis.

Equally important is the rationale for obtaining information from all the cumulated research. Research does no good to anyone if reasonable conclusions cannot be drawn from the research literature. Considering the vast number of studies addressing any particular research question today, it is crucial, not that more studies--differing only subtly from its predecessors in methodology, subjects, or design--be performed, but rather, that some integration, some synthesis, and general conclusions take place, thereby enabling researchers to refine existing theories and/or develop new ones. Such integration can best be accomplished by the same statistical treatments meta-analysis first proposed by Glass and his colleagues (Glass, McGaw, & Smith, 1981) that combines measures across studies using different independent and dependent variables, and that combines "good," well-designed studies with poorly designed studies (see below). Thus, the constructs of the independent and dependent variables vary across studies and are, therefore, not comparable (Hunter & Schmidt, 1990). Hunter and Schmidt offer two counterarguments for this criticism, a logical one and a methodological one. Logically, because a meta-analysis analyzes results (i.e., numbers) rather than studies, any set of numbers can be compared to any other without any logical contradiction. In fact, "the claim that only studies which are the same in <u>all</u> respects can be compared is self-contradictory; there is no need to compare them, since they would obviously have the same findings within statistical error" (Glass, 1977, p. 357). Methodologically, the presence of differences across study settings necessitates a meta-analysis because such a determination must be made on an empirical basis, rather than on a logical or semantic one.

<u>Mixing studies with different quality designs.</u> It is possible that results of meta-analyses are often uninterpretable because results from well-designed studies are mixed with those from less well-designed studies (Wolf, 1986). This criticism, however, can be handled empirically by coding the quality of the design employed in each study and examining whether the results differ between well-designed studies and poorly designed ones. It is possible that although there may be no effect size difference between the two types, poorly designed studies may have more effect size variation (Wolf, 1986).

Availability bias. It is possible that a meta-analysis will more likely include published studies that more often show statistical significance and larger effect sizes than unpublished studies; the latter are often not included in a meta-analysis (Hunter & Schmidt, 1990; Wolf, 1986) simply because they were not published and hence, not available. This is often called the "file drawer" problem (Rosenthal, 1979). This criticism is actually valid only for those studies that have never been publically reported. Otherwise, results in published books, dissertations, and theses, as well as those presented in unpublished papers at professional meetings, can be included in a meta-analysis and combined with the results from published articles. Alternatively, it is possible to estimate the number or additional studies with statistically nonsignificant results that would be necessary to reverse a conclusion drawn from a meta-analysis, thereby providing some estimate of the robustness and validity of the findings (Wolf, 1986).

Meta-analysis favors studies reporting more results. In a particular meta-analysis, if some studies report more results than others, then there will be a bias quantitatively favoring the results from the former over those of the latter, and making the results appear more reliable than they really are (Wolf, 1986). The counterargument has been made that it does not necessarily follow that just because one study reports one statistically significant result, all other results from that same study will be statistically significant as well. Additionally, whether different conclusions will be drawn for performing separate analyses for each different outcome (criterion or dependent variable; Kulik, 1983), lumping them all into the same analysis

(Glass, 1977), limiting oneself to a fixed number of results (say, two) from each study (Gilbert, McPeek, & Mosteller, 1977), or averaging all of the results from the same study (Wolf, 1986), is an empirical question and can only be answered by comparing these various methods in practice.

Correcting for attenuation. Hunter and Schmidt (1990) argued that meta-analysis should involve correcting for artifacts or attenuation. Rosenthal (1984), however, criticized safeguards designed to correct for artifacts or attenuation and proposed that corrected correlations or effect sizes are not as useful as the uncorrected values, because most investigators do not correct for measurement error anyway. Hunter and Schmidt reasoned that because different measures of a variable possess different levels of reliability (Hunter & Schmidt, 1990), failure to correct for measurement error and attenuation will inevitably lead to (a) systematic underestimation of actual relationships among constructs and (b) many different population values for the correlation between any two constructs.

Strengths of Meta-Analysis

Objectivity. Because meta-analysis does not make any prejudgments a priori, it is more likely to be more objective than the more traditional narrative literature reviews.

Integration. Because meta-analyses are quantitative descriptions of large bodies of research literature, the findings, when distilled down to some common metric or statistical summary, can often lead to stronger conclusions than can an impressionistic literature review.

Providing insights. In addition to its ability to highlight and answer the highly researched questions within a particular research domain, meta-analysis can also illuminate gaps in the extant knowledge base, thereby providing new directions for future research, such as interacting or mediating variables, and leading to the formulation of new hypotheses.

Variations (or Extensions) of Meta-Analysis

Study Effects Meta-Analysis

The study effects meta-analysis is considered an improvement over the Glassian-type meta-analysis (Hunter & Schmidt, 1990) and entails (a) including only one effect size from each study, thereby helping to ensure statistical independence within the meta-analysis, and (b) making some judgment about the methodological quality of the studies, including only the best ones (cf., Best Evidence Synthesis; Slavin, 1986). Overall, this method is a refinement, allowing clearer demarcation between independent and dependent variable constructs and permitting fine tuning of hypothesis formulation and subsequent experimental implementation.

Schmidt-Hunter Meta-Analysis Methods

The Schmidt-Hunter meta-analysis can be regarded primarily as an extension of the Glassian type of meta-analysis, but differs in only one major aspect: unlike the latter, the Schmidt-Hunter meta-analysis methods do not take the variance of observed effect sizes (S_{ES}^2) at face value. Instead, after determination of the mean effect size, the hypothesis that the variance of effect sizes is entirely due to statistical artifacts is tested. These artifacts include sampling error, differences in reliability of independent and dependent variable measures across studies, differences in range restriction across studies, differences in instrument validity across studies, and computational errors.

> Schmidt and Hunter (1990) developed methods of estimating and subtracting variance due to the first three of these five artifacts. Generally, if these three artifacts account for 75% or more of the observed S_{ES}^2, they conclude that the residual S_{ES}^2 is probably due to the remaining two artifacts and that true $S_{ES}^2 = 0$. (Hunter & Schmidt, 1990, p. 485)

One hallmark of the Schmidt-Hunter meta-analysis method is the correction for artifacts. When each study in a meta-analysis reports quantifiable artifact information, the validity coefficients are corrected individually, and the mean and variance of each corrected validity coefficient are computed. However, when studies report quantifiable artifacts only sporadically, an artifacts distribution must be used. For an artifacts distribution, the order is the reverse of correcting each artifact when each study provides artifact information. That is, first the mean and the variance of the uncorrected coefficients are computed. The mean and variance are then corrected to eliminate the effects of the various artifacts (Hunter, Schmidt, & Jackson, 1982).

<u>Homogeneity Test-Based Meta-Analysis</u>

The Homogeneity Test-Based meta-analysis was advanced primarily as an improvement over the Glassian-type meta-analysis and has as its major strength the ability to detect moderator variables. Hedges (1982) and others proposed the use of statistical tests as an aid in deciding whether study outcomes are more variable than would be expected from sampling error alone. If they are not, then there is no reason to search for moderators. Hedges (1982) extended the concept of homogeneity tests to develop a more general procedure for moderator analysis based on significance testing. Briefly, it entails partitioning the overall χ^2 statistic into the sum of within- and between-group chi-squares. The original set of effect sizes or correlations in the meta-analysis is then divided into successively smaller subgroups until the χ^2 statistics within the subgroups are statistically nonsignificant, indicating that sampling error can explain all the variation within the last set of subgroups.

The major problem of this test is that if there are a large number of moderators and a relatively small number of studies, then, eventually, the smallest subgroup could conceivably contain only a few (perhaps even one or two) studies. In this event, the original purpose of the meta-analysis--to quantitatively cumulate and synthesize across studies--is met only for each subgroup of studies, making generalizability difficult across moderator variables. Another potential problem with this method is based on (over)reliance on statistical significance testing--a practice that originally led to the conception of meta-analysis in the first place (Hunter & Schmidt, 1990).

Alternative Methods of Integration of Findings

Chi-Square Method

The chi-square method entails converting any inferential statistic (e.g., F, t, r) into an exact probability or p value using the appropriate statistical tables found in the appendices of any statistics text. These p values are then converted to chi-squares, that are then added to the degrees of freedom (df). Overall levels of statistical significance of the studies being combined are then determined by checking the χ^2 probability table for the summed values of chi-square and df (Gage, 1978). One problem with this method is that it is most useful if only a few statistical results are to be synthesized and if these results come from studies that are as similar to each other as possible (i.e., replications of each other). This method is a slight modification of the Fisher combined test (see below).

Cumulation of p Values

As the name indicates, this method cumulates significance levels across studies to produce an overall p-value for the set of studies as a whole (Hunter & Schmidt, 1990). A small value enables the researcher to conclude that there is an effect at work. As with the vote-counting method covered below, one major problem with this method is that although a combined statistically significant p-value may be reached, there is no mention of the magnitude of the effect. Rosenthal and Rubin (1978) later recognized this glaring limitation and compensated with the invocation of effect size analysis combined with p-values. Rosenthal's "file-drawer problem" can be ameliorated to some extent by the method of cumulating p-values across studies. Using Rosenthal's technique, it is possible to calculate the number of studies that have been "tucked away" (due to statistically nonsignificant results) because of an effect size of zero that would have to exist to bring the combined p-value down to the desired alpha level (e.g., 0.05, 0.01). Because this number usually turns out to be very large (on the order of 65,000; Rosenthal & Rubin, 1978), it is highly unlikely that there are 65,000 studies tucked away on any one topic.

<u>Traditional Voting-Counting Method</u>

This method entails classifying each statistical result into one of four categories, determined by statistical significance or not and positive or negative (Jackson, 1980). The vote-counting method allows the detection for statistically significant trends across studies, even if no results of any study itself are found to be statistically significant. However, there are several important and glaring limitations: (a) the method is biased in favor of large sample sizes which may show only small effect sizes (Hunter & Schmidt, 1990); (b) because the method makes no correction for differences in sample sizes across studies, it works best if all sample sizes are as equal as possible; (c) a large number of statistical results are necessary in order to detect any reliable trends (Gall, Borg, & Gall, 1996); (d) only the direction of the effect, rather than its magnitude, is determined (Hedges & Olkin, 1985); and (e) the results can lead to erroneous conclusions, especially in light of (b) and (d) above.

Two profound modifications to the vote-counting method have been devised in an attempt to address the most damaging limitations mentioned above, namely, (c) and (d). One is that which yields only a statistical significance level for a group of studies and the other provides a quantitative mean effect size estimate.

<u>Vote-counting methods that yield only significance levels.</u> These methods

basically employ a sign test to determine whether the observed frequencies of findings in the positive or negative directions depart significantly from the 50-50 split predicted under the null hypothesis (Rosenthal, 1978). However, these modifications are useful only when the null hypothesis is true and not when it is false.

> When the null hypothesis is not rejected in cumulative studies with high statistical power, this does provide an estimate of population effect size: zero. However, when the null hypothesis is false, the binomial or sign tests provide no estimate of effect size. This is a serious disadvantage. (Hunter & Schmidt, 1990, p. 473)

Vote-counting methods yielding estimates of effect sizes. If sample sizes are known for all studies, then one has only to estimate the effect size from either the proportion of positive results or from the proportion of positive statistically significant results. Hedges and Olkin (1985) have devised equations for calculating confidence intervals around these effect size estimates and will generally be wider (due to less use of information as a result of counting positive or significant positive results) from those resulting when effect sizes are determined individually for each study and then averaged, as proposed by Glass (1977). The limitation, then, of course, is that the vote-counting-based estimates of effect sizes should only be used when the information required to determine individual effect sizes from individual studies is not available (Hunter & Schmidt, 1990).

Combined Tests

Fisher combined test. The Fisher combined test is an alternate form of the chi-square test (above), also yielding a chi-square value using $\chi^2 = -2 \sum \log_e p$ (Wolf, 1986). Because of its exponential component, this method has superior asymptotic properties over other methods, although it suffers from several limitations. Mosteller and Bush (1954) noted long ago that it can yield results inconsistent with a simple sign test in situations where the majority of studies showed results in one direction with p values close to 0.50 (chance). In this situation, the sign test could easily reject the overall null hypothesis, while the Fisher procedure would be more conservative and would not. A more serious disadvantage of the Fisher test is its support for either outcome when two studies of equally and strongly significant results in opposite directions are obtained. Thus, suppose $p < 0.001$ favoring the experimental group and $p > 0.001$ favoring the control group occurs, both of which are combined for a $p < 0.01$ using the Fisher procedure. In this case, the Fisher combined test supports the significance of either outcome (Adcock, 1960).

Winer combined test. Winer (1971) proposed a procedure for combining independent tests that come directly from the sampling distribution of independent t-statistics in which the t-statistics associated with each test are summed and divided by the square root of the sum of the df associated with each t after each df has been divided by df - 2. Thus,

$$Z_c = \frac{\sum t}{\sqrt{\sum \left[\frac{df}{(df-2)} \right]}}$$

which is based on df/(df - 2) being the variance of a t-distribution, which is approximately normally distributed when $df > 10$. Thus, this procedure is inappropriate for tests based on $n < 10$.

Stouffer combined test. The Stouffer combined test, originally attributed to Stouffer and colleagues, is more fully described by Mosteller and Bush (1954) and by Rosenthal (1978). It bears a great deal of similarity to the Winer procedure for summing ts, except that the p-values are converted to zs, instead of to ts, and are then summed. The denominator then simplifies to the square root of the number of tests combined, and the complete expression takes the form $Z_c = \sum z/\sqrt{N}$ where N is the number of combined tests. This procedure is based on the sum of normal deviates being itself a normal deviate, with the variance equal to the number of observations summed. There are two major advantages to the Stouffer test: (a) the calculations are more straightforward than either the Fisher or Winer procedures, which necessitate logarithmic transformations and adjustments for df, respectively; and (b), although the results of the z procedure are slightly more powerful than those of the t procedure, the results of the two procedures are virtually identical (Wolf & Spies, 1981).

Best evidence synthesis. The best evidence synthesis method was proposed by Slavin (1986) in an attempt to improve upon the Glassian practice of lumping all studies, regardless of their quality, into the synthesis. As the name indicates, this method entails including only the most methodologically best studies in the meta-analysis; all other studies are not included in the meta-analysis.

Such best evidence syntheses focus on the "best evidence" in the field, including those studies having the highest internal and external validity, using clearly defined and well-justified inclusionary criteria. However, since it is virtually impossible to control for all the other confounding variables in a correlational study, best evidence synthesis must necessarily entail random assignment of subjects to groups. But in correlational studies, including these investigating the relationship between ratings of perceived exertion and each criterion variable (heart rate, blood lactate, oxygen uptake), assignment of subjects to groups is not usually done as a matter of cause. Typically, subjects of specified physical characteristics are recruited, screened, and then tested; there is rarely any use of a control group. Furthermore, many investigators have calculated validity coefficients among independent (criterion) variables (e.g., between heart rate and power output; Borg, Hassmen et al., 1987; Borg, Van den Burg et al., 1987). Thus, those relatively few studies that do use control groups and/or calculated validity coefficients among criterion variables (e.g., between heart rate and blood lactate) would be included under the best evidence inclusionary criteria. But because there are so few of them, a best evidence synthesis meta-analysis would hardly be warranted.

Summary

The purpose of this half of the literature review was to briefly review meta-analysis as a statistical technique for integrating findings across studies, its origin, its limitations and strengths, and several alternative integration techniques.

As will be evident in the Procedures section, the meta-analysis method of choice in this study will be the Hedges' Test of Homogeneity. The Borg RPE literature has many inconsistent results, much of it probably due to many different types of moderator variables. The issue of moderators will be addressed in this study. In addition, although this body of literature is large, it will not yield hundreds of studies, like many research domains in the educational or psychological fields. The number of usable studies in the perceived exertion literature will be much more modest.

Finally, because not every study reports quantifiable artifacts, an artifacts distribution to correct for bias will be applied to each study feature sample size-weighted validity coefficient. Such a correction will have the effect of increasing each mean validity coefficient, while decreasing the attendant sampling error variance.

CHAPTER III

PURPOSE AND OBJECTIVES

Because the body of literature on the Borg RPE Scale is immense (e.g., Noble & Robertson, 1996), and because there is a great deal of inconsistency regarding the validity of the Borg RPE Scale when it is related to various physiological measures, there is a dire need to provide some quantitative assessment, some global analysis of the extant literature. The relationship of ratings of perceived exertion with three major criterion variables will be determined: (a) heart rate, (b) blood lactate, and (c) oxygen uptake.

Presently, there is a wealth of inconsistent findings in establishing a positive relationship between ratings of perceived exertion and blood lactate levels. The disagreements are not so glaring between ratings of perceived exertion and heart rate and between ratings of perceived exertion and oxygen uptake, because most studies find a positive relationship. These inconsistencies must be resolved using established techniques of meta-analysis for correlations. For this purpose, this study intends to address the following questions:

1. What is the integrated correlation between heart rate measurements and ratings of perceived exertion? What study features (e.g., subjects' gender, the nature of the exercise/work task and exercise protocols used) might account for the variation of results across studies?

2. What is the integrated correlation between blood lactate and ratings of

perceived exertion? What study features (e.g., subjects' gender, the nature of the exercise/work task and exercise protocols used) might account for the variation of results across studies?

3. What is the integrated correlation between oxygen uptake and ratings of perceived exertion? What study features (e.g., subjects' gender, the nature of the exercise/work task and exercise protocols used) might account for the variation of results across studies?

CHAPTER IV

METHOD

From now on, heart rate measurements, blood lactate concentrations, and oxygen uptake will be collectively referred to as criterion variables. Furthermore, the variable "oxygen uptake" includes: percent maximal oxygen uptake (ml kg^{-1} min^{-1}), minute ventilation (ml min^{-1}), respiratory rate (breaths min^{-1}), and oxygen uptake (ml kg^{-1} min^{-1} or ml min^{-1}). Note that all four of these variables of oxygen uptake use comparable units (some unit volume per unit time), even though there is a distinction among them in the sports physiology/psychology literature. For the purposes of this meta-analysis, however, such a distinction was not be made and all four were lumped together under the heading of "oxygen uptake."

Many of the articles covered in the literature review have been included with other relevant articles and unpublished works (e.g., conference papers, dissertations, theses) from a search using the SPORT Discus database. This database has proven to be comprehensive in its listing of references dealing with perceived exertion, including those not normally appearing in the mainstream (conference papers, dissertations, theses). The PSYCHLIT, ERIC, and MEDLINE databases have not provided any references that SPORT Discus did not provide as well. Key words used included, either singly or in combination: perceived exertion, Borg, blood lactate and perceived exertion, heart rate and perceived exertion, oxygen uptake and perceived exertion, perceived work, and exercise intensity.

Study Features

Table 1 lists the major study features which may potentially affect the results of the studies, thus causing inconsistencies in the reported results. The potential moderating effect of these study features is the focus of this meta-analysis.

Coding of Study Features

Data selected for this meta-analysis were derived from studies that reported relationships between ratings of perceived exertion and one, two, or all three of the physiological criterion variables mentioned above under specific exercise conditions and using specific subject groups. Numerous study features were coded for the 60 studies in the final collection.

Subject Gender

Subject gender was included in this meta-analysis. Studies employing males, females, or both were included in this analysis. In cases in which a study did not differentiate between males and females, no distinction was made in the coding scheme either and was subsequently treated as missing data (although other study features of such a study were included).

Subject Fitness/Activity Level

The aerobic fitness levels of subjects were included in the meta-analysis when the levels were measured by oxygen uptake. The subjects' fitness levels were

Table 1

Features of Studies

General study feature	Specific study features
Subject gender	(1) Male (2) Female
Subject fitness/activity level	(1) Sedentary (2) Healthy - nonactive (3) Healthy - active (4) Highly fit
Type of RPE Scale	(1) 15-point (2) 21-point (3) 9-point (4) Category-ratio (10 or 20)
Exercise type or work mode	(1) Bicycle ergometer (2) Treadmill (3) Track running (or Treadmill running) (4) Arm cranking (5) Swimming
Exercise protocol	(1) Progressive continuous (2) Progressive intermittent (3) Random intermittent (4) One-level--maximal exertion (5) One-level--submaximal exertion
Quality of study	(1) Excellent--all necessary design, controls incorporated (2) Good--some, but not all, controls incorporated (3) Poor--fairly lacking in proper experimental procedures

operationally defined as a reference to their particular lifestyle and were categorized as: (a) sedentary, (b) healthy, but typically inactive, (c) active, and (d) highly fit. Maximum oxygen uptake is a routine physiological measure that indicates how well an individual aerobically adapts to the increased metabolic demands of exercise. That is, oxygen uptake is positively related to activity level. The categorization by activity level (reported in each article) is then determined by a combination of the oxygen uptake measure and a subjective appraisal of the subjects' fitness level.

Type of RPE Scale Used

Four types of RPE scales were considered for this meta-analysis: (a) 15-point, (b) 21-point, (c) 9-point, and (d) the category-ratio scale.

Exercise Type

The five most prevalent forms of exercise in the perceived exertion literature are (a) bicycle ergometer, (b) treadmill, (c) track running, (d) arm cranking, and (e) swimming or deep water running. In the final analysis, however, all studies using arm cranking were excluded because they failed to report Pearson *r*s in their results.

Exercise Protocol

Five exercise protocols were considered. The (a) progressive continuous and (b) progressive intermittent protocols start with low levels of exercise and progress to more strenuous levels of exercise. In the progressive continuous protocol, subjects exercise without any rest between workloads. In the progressive intermittent protocol, subjects exercise with rest periods between each workload. The (c) random intermittent protocol has a random, rather than a progressive, order of exercise levels with rest provided between each workload. The (d) maximal exertion protocol remains at one exercise level at which the subject is to continue to exercise until (s)he becomes physically exhausted. And the (e) submaximal exertion protocol also remains at one exercise level at which the subject exercises in a physical steady state for a specified period of time.

Study Quality

The quality of the study is a characteristic that would obviously not be reported in any study itself, but rather is determined via the coding scheme for this meta-analysis: (a) a rating of "excellent" indicates that all necessary and sound designs (e.g., randomization, delineation of variables, adequate description) and controls were incorporated, and explained; (b) "good" means that most, but not all, experimental procedures were practiced and explained. The number of such transgressions was set at ≤ 2; and (c) a rating of "poor" meant that more than two such experimental design transgressions occurred in the study.

Criteria for Acceptance of a Study into the Meta-Analysis

Any study exploring the relationship between ratings of perceived exertion and any of the three criterion variables will be included in this meta-analysis. Specifically, any study reporting Pearson rs or any statistic that can be converted to a Pearson r

(see below) were included in this meta-analysis. The study quality is part of the category/feature coding listed in Table 2, and will be one variable used to differentiate studies well-conducted studies from more poorly conducted ones.

Coding Scheme

Pertinent information to be coded from each article is listed in Table 1. In the Appendix (Tables A.1, A.2, and A.3), the summaries of characteristics are coded for each article dealing with the relationship between ratings of perceived exertion and heart rate, blood lactate, and oxygen uptake, respectively.

As can be seen in Tables A.1, A.2, and A.3, some of the studies include missing values. For example, Brown et al. (1996b; Table A.1) did not indicate the gender of their subjects. In such cases, "gender" for this study was simply counted as (a) missing value(s).

Statistics

Studies reporting Fs, ts, regression slopes (if units across studies are comparable), and so forth were converted to Pearson rs as prescribed by Glass (1977, p.374). For example,

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_x^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad (1)$$

$$r_{pb} = \sqrt{\frac{t^2}{t^2 + (n_1 + n_2 - 2)}} \qquad (2)$$

$$r_{xy} = r_{pb}\frac{\sqrt{n_1 n_2}}{un} \qquad (3)$$

$$F = \frac{MS_{bet}}{MS_{within}} \; for \; j = 2 \; groups$$

$$F = t^2 \qquad (4)$$

where r_{pb} is the point-biserial correlation, u is the ordinate (height) of the unit normal distribution at the point above which lies 100 (n_1/n) % of the area under the curve, and $n = n_1 + n_0$ = the numbers of 1s and 0s on y, respectively. Because biases do not necessarily exist in some studies that report more statistically significant results than others, all reported correlations (or other statistics that have been converted to rs) were used (Glass, 1977).

Fisher's z-Transformation

Fisher's z-transformation was first performed on each study validity coefficient r

$$Z_r = \tanh^{-1} r \tag{5}$$

and its accompanying standard error:

$$\sigma_Z = \frac{1}{\sqrt{n - 3}} \tag{6}$$

From the nearly normal sampling distribution of zs, the resultant z_r from equation (5) corresponds to the mean, which, in turn, corresponds to ρ; that is, z_ρ

Testing Homogeneity of Correlations Across Studies for Each Criterion Variable

After all study test statistics (e.g., t, F) have been transformed to Pearson rs and appropriately transformed to Fisher's zs, the first statistical test was the test for homogeneity of correlations across studies because Borg Scale studies often use relatively small sample sizes (Table 1). To this end, therefore, the Hedges' Test for

Homogeneity of Correlations was used. The analyses conducted in this study was based on Hedges' (1981, 1982) tests for fitting categorical models to effect sizes. Correlations representing the relationship with each of the three perceived exertion-criterion variables were analyzed separately to test whether the data are reasonably consistent with the model of a single underlying population correlation. Hedges' Test of Homogeneity, H_T, was used:

$$H_T = \sum (n_i - 3)(z_i - \bar{z})^2 \tag{7}$$

where z_i represents an individual Fisher's z-transformed correlation and the mean of the z_is represents the average correlation weighted by sample size. The larger the deviation ($z_i - \bar{z}$), the more inconsistency exists among studies. Each n_i is the sample size for the specific correlation where ($n_i - 3$) is the variance of each z. The Hedges' Test of Homogeneity, which is analogous to the analysis of variance (ANOVA), was used to examine homogeneity of correlations within groups (within each criterion variable). The ANOVA-like comparisons were based on the relationship: $H_B = H_T - H_W$, where H_T represents the total homogeneity value across the correlations representing the relationship with each RPE-criterion variable. H_W represents the total within-group homogeneity; that is, homogeneity within each criterion variable. And H_B represents the difference between the total value of the correlations representing the relationship with each perceived exertion-criterion variable and the total within-study group (criterion variable) H_W value. The H_T statistic is an indication of whether the sample correlations seem more varied than would be expected on the basis of sampling variability and is compared to a χ^2 distribution with $\underline{k} - 1$ $\underline{df}$ (where $\underline{k}$ is the number of correlations). If the H_T value for any set of correlations representing the relationship between ratings of perceived exertion and the criterion variable was less than the χ^2 value, it was not statistically significant and the correlations are considered homogeneous, indicating that the correlations representing this relationship with that particular perceived exertion-criterion variable were similar regardless of the studies from which they were drawn, such as those using one exercise type or subject gender as opposed to another.

If, however, the Test of Homogeneity is statistically significant, then the correlations representing the relationship with each of the three perceived exertion-criterion variables were not consistent and the average correlation value cannot be generalized across studies. In this event, the correlations of ratings of perceived exertion with each of the three criterion variables were grouped by each of the study features listed in Table 1. The extent of this inconsistency of the validity coefficients among study features (e.g., exercise type: bicycle ergometer, treadmill, track running, and swimming) was then determined using a one-way ANOVA, followed by a Tukey's multiple comparison test.

Averaging Correlations

For each criterion variable correlated with ratings of perceived exertion, grouping studies according to the types of studies used (e.g., male vs. females, sedentary vs. active), it was assumed that the population correlation is constant over studies <u>within</u> a particular gender or occupation group. In such cases, the best estimate of the population correlation is a weighted mean in which each correlation is weighted by the sample size of that study (Hunter & Schmidt, 1990). Two types of mean sample size-weighted correlations were calculated: (a) a weighted $\bar{r}$ derived from $\Sigma(n_i\, r_i)/N_i$. For this particular mean correlation, there was no $\bar{r}$ - to - $\bar{z}$ - to - $\bar{r}$ transformation. This $\bar{r}$ is denoted M_1. And (b) a weighted $\bar{r}$ derived from the mean z, $\Sigma(n_i\, z_i)/N_i$, where z_i is the Fisher's z transformed r. The resultant $\bar{z}$ was then back-transformed to its corresponding $\bar{r}$. This $\bar{r}$ is denoted M_2. Likewise, the corresponding variance across studies is the frequency weighted average squared error (Hunter & Schmidt, 1990):

$$s_r^2 = \mathrm{Var}(r) = [\Sigma[N_i\,(r_i - \bar{r})^2]]/\Sigma N_i \qquad (8)$$

<u>Searching for moderator variables.</u> For each perceived exertion-criterion variable relationship, the next step was to calculate a one-way ANOVA for each study feature (e.g., exercise type) and then a Tukey multiple-comparisons test to determine which of the four types of exercises (e.g., bicycle ergometer, treadmill, track running, and swimming) are statistically significantly different from each other.

Finally, the last step for each perceived exertion-criterion variable relationship was to corroborate or refute the homogeneity test. To test for statistically significant interactions and, therefore, the lack of generalizability across studies (Glass & Hopkins, 1996, p. 485), or the converse, all 15 possible pairwise combinations between two study characteristics (e.g, fitness level by RPE Scale) were each examined in a two-way ANOVA.

Correcting for Artifacts--The Use of Artifact Distributions

Hunter and Schmidt (1990; Hunter et al., 1982) proposed the use of artifact distributions when studies report quantifiable artifacts only sporadically. When validity coefficients are corrected individually, the mean and variance of each corrected coefficient is computed. For artifact distributions, however, the order is reversed. First, the mean and the variance of the uncorrected coefficients are computed. This mean and variance are then corrected to eliminate the effects of the various artifacts (Hunter et al., 1982). Table 2 lists and briefly explains many of the artifacts which Hunter and Schmidt (1990, p. 45) provided, but that are most applicable to educational and occupational psychological fields of study.

Partly due to the nature of the Borg RPE Scale itself (one item), to the physiological criterion variables (e.g., heart rate, blood lactate oxygen uptake, power output), to the types of subjects who volunteer for such studies, and to the designs typically used, only two artifacts (nos. 1 and 9) are applicable to the kinds of studies typically and most commonly implemented in the ratings of perceived exertion literature. Further, within the ratings of perceived exertion literature, qualitative artifact information is actually often provided (Ekblom & Goldbarg, 1971), but quantitative information across studies is quite sporadic. Much of the artifact information is only qualitatively provided (described) due to its inherent presence in

Table 2

Study Artifacts That Cause the Study Validity Coefficient to Differ from the True (Population) Correlation and the Corrections Associated with These Artifacts

---

1. Sampling error will cause study validity to randomly vary from the population value.

Correction: The variance of population correlations $\sigma_\rho^2 = \sigma_r^2 - \sigma_e^2$ where $\sigma_e^2 = \frac{(1-\bar{r}^2)^2 k}{T}$ where k is the number of correlations and $T = \Sigma N_i$ is the total sample size.

Application to Perceived Exertion Studies: Applicable to perceived exertion studies.

2. Error of Measurement in the Dependent Variable: Study validity will be systematically lower than true validity to the extent that Borg Scale ratings of perceived exertion are measured with random error --Reliability of the dependent variable (Borg Scale) (e.g., test-retest).

Correction: Same as Sampling Error: $\sigma_r^2 = s_r^2$ (Equation 8). For reliability, $r - r_o/(\sqrt{r_{xx}}\sqrt{r_{yy}})$ where r_o is the study sample uncorrected correlation and r_{xx} and r_{yy} are the reliability coefficients of the dependent (RPE Scale) and independent (heart rate, blood lactate, or oxygen uptake) variables, respectively.

Application to Perceived Exertion Studies: Not applicable to perceived exertion studies. Values of each of these three physiological criterion variables will depend on subjects' state of rest or exercise at the time the measurement is made. Thus, reliability does not apply to physiological measures inasmuch as, say, heart rate, is considered reliable.

3. Error of Measurement in the Independent Variable--Reliability of the Independent variable. As in no. 2 preceding.

Correction: $r - r_o/(\sqrt{r_{xx}}\sqrt{r_{yy}})$ where r_o is the study sample uncorrected correlation and r_{xx} and r_{yy} are the reliability coefficients of the dependent (RPE Scale) and independent (heart rate, blood lactate, or oxygen uptake) variables, respectively.

Application to Perceived Exertion Studies: Not applicable to perceived exertion studies. Values of each of these three physiological criterion variables will depend on subjects' state of rest or exercise at the time the measurement is made. Thus, reliability does not apply to physiological measures inasmuch as, say, heart rate, is considered reliable.

4. Range variation in the dependent variable will be systematically lower than true validity to the extent that volunteers will exhibit response biases, needs or desires to impress others (such as other participants, the experimenter, etc.), and basically be less than honest in rating themselves on the Scale. Other factors influencing responses will be motivation (fun vs. competition), desire, and so forth laboratory versus natural setting.

Correction: Cannot be corrected.

5. Imperfect Validity of the Independent Variable (heart rate, oxygen uptake blood lactate levels, power output, speed, etc.) In which an independent variable is correlated with another independent variable (e.g., heart rate and oxygen uptake or oxygen uptake and treadmill speed).

Correction: $r = r_o/ab$ where a and b are the correlations between independent variables (e.g., a = the correlation between heart rate and oxygen uptake and b = the correlation between oxygen uptake and power output).

(table continues)

Application to Perceived Exertion Studies: Not applicable to perceived exertion studies. For physiological measures, validity of their interpretation is limited to the validity of the technology, instruments, and/or laboratory procedures designed to measure them.

6. Imperfect Construct Validity of Perceived Exertion: Study validity will differ from true validity if the criterion is somehow deficient, contaminated, or otherwise undermined (e.g., subjects receiving inadequate instructions in the proper use of the Borg Scale).

Correction: Corrected sample correlation is $r_i = r_0 / \bar{r}$, where r_0 is the study sample uncorrected correlation. Also, when provided, correlations of RPEs with some other independent variable will be the divisor to correct the sample r_o (e.g., $r_{RPE,\ speed}$); that is, $r = r_o / r_{RPE,\ speed}$.

Application to Perceived Exertion Studies: The validity of such independent variables (e.g., speed), again, is limited to the validity of the technology, instruments, and/or laboratory procedures designed to measure them.

7. Reporting on transcriptional error.

Correction: Cannot be corrected.

8. Variance due to Extraneous Factors: Study validity will be systematically lower than true validity if volunteers differ in physical fitness condition, or daily health habits (e.g., smoking, obesity, drinking), or if the exercise/work task is performed and the Scale administered under different conditions, time of day, et cetera.

Correction: Where and when information on extraneous factors is given, the corrected correlation is $r = r_0 / \sqrt{(1 - \rho_{EB}^2)}$ where ρ_{EB} is the correlation between the extraneous factor and the Borg RPE scores.

Application to Perceived Exertion Studies: Not applicable to perceived exertion studies. Such extraneous factors are myriad and so varied that it may not be generalizable across studies and individuals, especially when studies report these extraneous factors only sporadically, thereby necessitating the use of artifact distributions. For example, the measures obtained from obese subjects may not be applicable to those obtained from other kinds of subjects.

9. Bias in the Correlation: Hunter and Schmidt (1990) point out that the presence of statistical bias in the sample correlation as an estimate of the population correlation is usually quite trivial in magnitude and is, therefore, rarely worth the trouble to correct for it. "However, we provide the computations to check the size of the bias in any given application. The impact of bias is systematic and can be captured to a close approximation by an attenuation multiplier" (p. 141).

Correction: The correction takes the general form $r = r_o / [(2N - 2)/(2N - 1)]$.

Application to Perceived Exertion Studies: This is the only other correction that applies to perceived exertion studies. Systematic application of this multiplier will naturally decrease the mean validity coefficients.

Adapted from Hunter and Schmidt (1990, p. 45).

---

this type of research. Miller et al. (1985) stated that "on any given day, one's RPE and heart rate (and oxygen values) may fluctuate with exercise as a result of physical, social and emotional factors" (pp. 193-194).

Study Effects Meta-Analysis

A study effects meta-analysis was performed in which the validity coefficient between ratings of perceived exertion and each of the three criterion variables is examined on a by-study basis (Bangert-Drowns, 1986). Each study contributed only one validity coefficient, whether that coefficient was a single correlation or the mean of more than one coefficient.

CHAPTER V

RESULTS

Sixty studies were found to meet the inclusionary criteria stated in the preceding chapter; that is, these 60 studies revealed Pearson *r*s between ratings of perceived exertion and one or more of the criterion variables being examined in this meta-analysis. Out of these 60 studies, 161 validity coefficients were found between ratings of perceived exertion and heart rate, 36 validity coefficients were found between ratings of perceived exertion and blood lactate, and 74 validity coefficients were found between ratings of perceived exertion and oxygen uptake. Together, therefore, these 60 studies generated 271 *r*s.

Test of Homogeneity

As can be seen in Table 3, all three criterion variables are statistically significant, indicating that their validity coefficients with ratings of perceived exertion are heterogeneous. This means that the correlations between ratings of perceived exertion and each of the three criterion variables cannot be generalized across studies, and are the result of moderating features contributing greater variability than would be expected from sampling variability alone.

The six moderator variables (Table 1) explored in this meta-analysis, therefore, were examined separately for each of the three criterion variables first by determining the mean validity coefficients by each study feature and then by using a one-way

Table 3

Homogeneity Test for Correlations Between Ratings of Perceived Exertion and the Three Criterion Variables

Study feature	H_T	H_W	H_B	${}_{.05}\chi^2_{df}$	$\underline{M}^a$
Heart rate	2,568.51*	1,614.42	954.07	190.23	.616
Blood lactate	337.70*	231.22	106.48	49.80	.591
Oxygen uptake	613.73*	246.68	367.06	93.95	.532

[a]Correlations have been back-transformed from their mean sample size-weighted Fisher's $\underline{z}$s. *$\underline{p} < .05$.

ANOVA and subsequent multiple comparisons test. Both types of analyses (mean validity coefficients and ANOVA/multiple comparisons tests) will reveal which study features (e.g., treadmill, bicycle ergometer subsumed under exercise type) will yield mean validity coefficients that are higher or lower than the rest of the study features, and if so, if they are statistically significantly different from the rest of the study features.

These study features yielding mean validity coefficients that are statistically significantly different from the rest within a particular feature group (e.g., exercise type) indicate that they are moderating variables and are responsible for the observed heterogeneity within each perceived exertion-criterion variable relationship. That is, they contribute more variability than can be accounted for by sampling variability alone.

Mean Validity Coefficients

Because the $\underline{\bar{z}}$ - to - $\underline{\bar{r}}$ back-transformations depend on means, rather than on individual validity coefficients, ANOVA and the subsequent Tukey multiple comparisons tests were calculated only for the n-weighted mean validity coefficient (M_1). To make the Tukey's multiple comparison test meaningful, Tables A.4, A.5, and A.6 provide the descriptive statistics of the weighted validity coefficients (M_1) for heart rate, blood lactate, and oxygen uptake, respectively.

Heart Rate

Reported in Tables 4, 5, 6, 7, 8, and 9 are the mean (M) sample-size-weighted validity coefficients (M_1 and M_2), the number of correlations (k), and number of subjects (n) used to derive each mean validity coefficient for each of the study features. The two types of mean coefficients, M_1 and M_2, for heart rate are .532 and .616, respectively. As can be seen from Tables 4-9, all mean correlations for all six study characteristics and their respective study features are in the low to high range (approximately .300 [Table 9] to .985 [Table 8]). Of the six study characteristics, only the features within fitness level (Table 5), exercise type (Table 7), exercise protocol (Table 8), and study quality (Table 9) are statistically significantly different from each other. Specifically, Tukey's multiple comparison test reveals the following for the

Table 4

Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Heart Rate by Gender

Feature	n-weighted M_1	n-weighted M_2	k[a]	n[b]
Overall	.531	.616	143	3,570
Males	.525	.633	94	2,078
Females	.538	.591	49	1,492

ANOVA results

Source	df	F	p	η^2 [c]
Gender	1,141	1.44	> .05	.0101

[a] k is the number of correlations, Overall $k < 161$ due to missing values.
[b] n is the number of subjects.
[c] η^2 is eta squared, which is defined as (sum of squares $_{\text{main effect}}$)/(sum of squares $_{\text{total}}$).

Table 5

Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Heart Rate by Fitness Level

Feature	n-weighted M_1	n-weighted M_2	k[a]	n[b]
Overall	.545	.650	123	2,665
Sedentary	.379 (A)[c]	.534	19	231
Healthy-inactive	.521 (B)	.574	43	1,361
Active	.593 (B)	.714	38	758
Highly fit	.660 (B)	.815	23	315

ANOVA results

Source	df	F	p	η^2 [d]
Fitness level	3, 119	5.34	< .05	.119

[a] k is the number of correlations. Overall k < 161 due to missing values.
[b] n is the number of subjects.
[c] (A) is statistically significantly different from (B).
[d] η^2 is eta squared, which is defined as: (sum of squares $_{\text{main effect}}$)/(sum of squares $_{\text{total}}$).

Table 6

Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Heart Rate by RPE Scale

Feature	n-weighted M_1	n-weighted M_2	k[a]	n[b]
Overall	.532	.616	161	3,957
15-point	.535	.626	142	3,441
21-point	.605	.612	6	154
9-point	.549	.559	3	115
Category ratio	.450	.505	10	247

ANOVA results

Source	df	F	p	η^2[c]
RPE Scale	3, 157	1.23	> .05	.023

[a] k is the number of correlations.
[b] n is the number of subjects.
[c] η^2 is eta squared, which is defined as: (sum of squares $_{\text{main effect}}$)/(sum of squares $_{\text{total}}$).

Table 7

Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Heart Rate by Exercise Type

Feature	n-weighted M_1	n-weighted M_2	k[a]	n[b]
Overall	.533	.618	160	3,933
Bicycle ergometer	.577	.655	108	2,064
Treadmill	.482 (B)[c]	.556	37	1,598
Track running	.332 (B)	.499	4	166
Swimming	.778 (A)	.834	11	105

ANOVA results

Source	df	F	p	η^2 [d]
Exercise type	3, 156	4.63	< .05	.300

[a] k is the number of correlations. Overall k < 161 due to missing values.
[b] n is the number of subjects.
[c] (A) is statistically significantly different from (B).
[d] η^2 is eta squared, which is defined as: (sum of squares $_{\text{main effect}}$)/(sum of squares $_{\text{total}}$).

Table 8

Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Heart Rate by Exercise Protocol

Feature	n-weighted M_1	n-weighted M_2	k[a]	n[b]
Overall	.532	.616	161	3,957
Progressive continuous	.583 (B)[c]	.637	58	1,506
Progressive intermittent	.698 (B) (C)	.737	18	321
Random intermittent	.450 (B) (D)	.540	18	262
One-level maximal exertion	.841 (A)	.985	10	87
One-level submaximal exertion	.457 (B) (D)	.525	57	1,781

ANOVA results

Source	df	F	p	η^2 [d]
Exercise Protocol	4, 156	11.64	< .05	.230

[a] k is the number of correlations.

[b] n is the number of subjects.

[c] (A) is statistically significantly different from (B) and (C) is statistically significantly different from (D).

[d] η^2 is eta squared, which is defined as: (sum of squares $_{\text{main effect}}$)/(sum of squares $_{\text{total}}$).

Table 9

Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Heart Rate by Study Quality

Feature	n-weighted M_1	n-weighted M_2	k[a]	n[b]
Overall	.528	.612	152	3,875
Excellent	.530 (B)[c]	.647	112	2,051
Good	.587 (B)	.625	32	1,425
Poor	.304 (A)	.328	8	399

ANOVA results

Source	df	F	p	η^2 [d]
Study quality	2, 149	3.38	< .05	.0433

[a] k is the number of correlations. Overall k < 161 due to missing values.
[b] n is the number of subjects.
[c] (A) is statistically significantly different from (B).
[d] η^2 is eta squared, which is defined as: (sum of squares $_{\text{main effect}}$)/(sum of squares $_{\text{total}}$).

mean weighted validity coefficients between ratings of perceived exertion and heart rate: (a) for fitness level, for sedentary individuals are statistically significantly different from those of the other three fitness types (Table 5); (b) for exercise type, swimming is statistically significantly different from the treadmill and track running (Table 7); (c) for exercise protocol, the maximal exertion protocol is statistically significantly different from the other four protocols (Table 8); and (iii) for study quality, "excellent" studies and "good" studies are statistically significantly different from those coded as "poor" (Table 9).

Blood Lactate

Reported in Tables 10, 11, 12, 13, 14, and 15 are the mean (M) sample-size-weighted validity coefficients, the number of correlations (k), and number of subjects (n) used to derive each mean correlation between ratings of perceived exertion and blood lactate.

The two types of mean correlations, M_1 and M_2 for blood lactate are .530 and .591, respectively. As can be seen from Tables 10-15, all validity coefficients for all six study characteristics and their respective study features are in the mid to high range (ca. .412 [Table 12] to .840 [Table 10]). Four out of the six study features, gender (Table 10), RPE Scale (Table 12), exercise type (Table 13), and exercise protocol (Table 14) show between-group statistically significant differences. Specifically, Tukey's multiple comparison test reveals that: (a) for gender, the mean weighted validity coefficients for males are statistically significantly different from those for

Table 10

Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Blood Lactate by Gender

Feature	n-weighted M_1	n-weighted M_2	k[a]	n[b]
Overall	.530	.591	36	661
Males	.463	.489	23	521
Females	.778	.839	13	140

ANOVA results

Source	df	F	p	η^2[c]
Gender	1, 34	15.69	< .05	.3157

[a] k is the number of correlations.
[b] n is the number of subjects.
[c] η^2 is eta squared, which is defined as: (sum of squares $_{\text{main effect}}$)/(sum of squares $_{\text{total}}$).

Table 11

Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Blood Lactate by Fitness Level

Feature	n-weighted M_1	n-weighted M_2	k[a]	n[b]
Overall	.532	.595	32	633
Healthy-inactive	.674	.802	8	103
Active	.499	.535	23	503
Highly fit	.610	.610	1	27

ANOVA results

Source	df	F	p	η^2 [c]
Fitness level	2, 29	1.63	> .05	.1011

[a] k is the number of correlations. Overall $k < 36$ due to missing values.
[b] n is the number of subjects.
[c] η^2 is eta squared, which is defined as: (sum of squares $_{main\ effect}$)/(sum of squares $_{total}$).

Table 12

Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Blood Lactate by RPE Scale

Feature	n-weighted M_1	n-weighted M_2	k[a]	n[b]
Overall	.530	.591	36	661
15-point	.721 (A)[c]	.785	22	230
9-point	.457 (B)	.457	3	150
Category-ratio	.412 (B)	.438	11	281

ANOVA Results

Source	df	F	p	η^2 [d]
RPE Scale	2, 33	12.47	< .05	.430

[a] k is the number of correlations.
[b] n is the number of subjects.
[c] (A) is statistically significantly different from (B).
[d] η^2 is eta squared, which is defined as: (sum of squares $_{\text{main effect}}$)/(sum of squares $_{\text{total}}$).

Table 13

Mean Sample-Size-Weighted Validity Coefficients for Blood Lactate by Exercise Type

Feature	n-weighted M_1	n-weighted M_2	k[a]	n[b]
Overall	.530	.591	36	661
Bicycle ergometer	.477 (A)[c]	.527	24	515
Treadmill	.712 (B)	.768	10	126
Swimming	.755 (B)	.756	2	20

ANOVA results

Source	df	F	p	η^2 [d]
Exercise type	2, 33	3.41	< .05	.263

[a] k is the number of correlations. Overall k < 36 due to missing values.
[b] n is the number of subjects.
[c] (A) is statistically significantly different from (B).
[d] η^2 is eta squared, which is defined as: (sum of squares $_{\text{main effect}}$)/(sum of squares $_{\text{total}}$).

Table 14

Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Blood Lactate by Exercise Protocol

Feature	n-weighted M_1	n-weighted M_2	k[a]	n[b]
Overall	.530	.591	36	661
Progressive continuous	.494 (B)[c]	.595	17	306
Progressive intermittent	.735 (A)	.746	10	103
Random intermittent	.457 (B)	.457	3	150
One-level maximal exertion	.480 (B)	.482	2	44
One-level submaximal exertion	.582 (B)	.630	4	58

ANOVA results

Source	df	F	p	η^2 [d]
Exercise protocol	4, 31	11.24	< .05	.2375

[a] k is the number of correlations.
[b] n is the number of subjects.
[c] (A) is statistically significantly different from (B).
[d] η^2 is eta squared, which is defined as: (sum of squares $_{\text{main effect}}$)/(sum of squares $_{\text{total}}$).

Table 15

Mean Sample-Size-Weighted Validity Coefficient and ANOVA for Blood Lactate by Study Quality

Feature	n-weighted M_1	n-weighted M_2	k[a]	n[b]
Overall	.530	.591	36	661
Excellent	.523	.585	31	602
Good	.600	.648	5	59

ANOVA results

Source	df	F	p	η^2 [c]
Study quality	1, 34	.03	> .05	.001

[a] k is the number of correlations.
[b] n is the number of subjects.
[c] (A) is statistically significantly different from (B).
[d] η^2 is eta squared, which is defined as: (sum of squares $_{\text{main effect}}$)/(sum of squares $_{\text{total}}$).

females, whose responses were quite a bit higher (Table 10), (b) for RPE scale usage, the 15-point scale is statistically significantly different from those of both the 9-point scale and the category-ratio scale (Table 12), (c) for exercise type, use of the bicycle ergometer is statistically significantly different from both treadmill and swimming (Table 13), and (d) for exercise protocol, the progressive intermittent protocol is statistically significantly different from each of the other four protocols (Table 14).

Oxygen Uptake

Reported in Tables 16, 17, 18, 19, 20, and 21 are the mean (M) sample-size-weighted validity coefficients, the number of correlations (k), and number of subjects (n) used to derive each mean validity coefficient between ratings of perceived exertion and oxygen uptake.

The two types of mean validity coefficients, M_1 and M_2, for oxygen uptake are .427 and .532, respectively. As can be seen from Tables 16-21, unlike for heart rate and blood lactate, all mean validity coefficients for all six study characteristics and their respective study features are in the low to high range (ca. -.055 [Table 20] to .858 [Table 19]). Of the six study characteristics, only the feature of RPE Scale does not show any between-group statistically significant differences (Table 18). Specifically, Tukey's multiple comparison test reveals that in the sample size-weighted validity coefficients between ratings of perceived exertion and oxygen uptake, males are statistically significantly different from females (Table 16), active individuals are statistically significantly different from healthy-inactive and highly fit individuals (Table

Table 16

Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Oxygen Uptake by Gender

Feature	n-weighted M_1	n-weighted M_2	k[a]	n[b]
Overall	.425	.538	63	944
Males	.577	.650	55	628
Females	.123	.252	8	316

ANOVA Results

Source	df	F	p	η^2 [c]
Gender	1, 61	4.79	< .05	.2128

[a] k is the number of correlations. Overall $k < 74$ due to missing values.
[b] n is the number of subjects.
[c] η^2 is eta squared, which is defined as: (sum of squares $_{\text{main effect}}$)/(sum of squares $_{\text{total}}$).

Table 17

Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Oxygen Uptake by Fitness Level

Feature	n-weighted M_1	n-weighted M_2	k[a]	n[b]
Overall	.376	.476	44	996
Healthy-inactive	.509 (B)[c]	.547	24	454
Active	.241 (A)	.396	18	505
Highly fit	.582 (B)	.594	2	37

ANOVA results

Source	df	F	p	η^2 [d]
Fitness Level	2, 41	8.34	< .05	.2630

[a] k is the number of correlations. Overall k < 74 due to missing values.
[b] n is the number of subjects.
[c] (A) is statistically significantly different from (B).
[d] η^2 is eta squared, which is defined as: (sum of squares $_{main\ effect}$)/(sum of squares $_{total}$).

Table 18

Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Oxygen Uptake by RPE Scale

Feature	n-weighted M_1	n-weighted M_2	k[a]	n[b]
Overall	.427	.532	74	1,170
15-point	.408	.527	69	1,012
Category-ratio	.549	.560	5	158

ANOVA results

Source	df	F	p	η^2 [c]
RPE Scale	1, 72	.14	> .05	.002

[a] k is the number of correlations.
[b] n is the number of subjects.
[c] η^2 is eta squared, which is defined as: (sum of squares $_{\text{main effect}}$)/(sum of squares $_{\text{total}}$).

Table 19

Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Oxygen Uptake by Exercise Type

Feature	n-weighted M_1	n-weighted M_2	k[a]	n[b]
Overall	.427	.532	74	1,170
Bicycle ergometer	.432 (B)[c]	.542	48	638
Treadmill	.331 (B)	.356	18	441
Swimming	.858 (A)	.915	8	91

ANOVA results

Source	df	F	p	η^2 [d]
Exercise type	2, 71	10.69	< .05	.2315

[a] k is the number of correlations.
[b] n is the number of subjects.
[c] (A) is statistically significantly different from (B).
[d] η^2 is eta squared, which is defined as: (sum of squares $_{\text{main effect}}$)/(sum of squares $_{\text{total}}$).

Table 20

Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Oxygen Uptake by Exercise Protocol

Feature	n-weighted M_1	n-weighted M_2	k[a]	n[b]
Overall	.427	.532	74	1,170
Progressive continuous	.413 (B) (C)[c]	.502	31	688
Progressive intermittent	.565 (B)	.770	14	111
Random intermittent	.750 (B) (D)	.750	1	10
One-level maximal exertion	-.055 (A)	.059	4	80
One-level submaximal exertion	.532 (B)	.580	24	281

ANOVA results

Source	df	F	p	η^2 [d]
Exercise protocol	4, 69	6.31	< .05	.279

[a] k is the number of correlations.

[b] n is the number of subjects.

[c] (A) is statistically significantly different from (B) and (C) is statistically significantly different from (D).

[d] η^2 is eta squared, which is defined as: (sum of squares $_{\text{main effect}}$)/(sum of squares $_{\text{total}}$).

Table 21

Mean Sample-Size-Weighted Validity Coefficients and ANOVA for Oxygen Uptake by Study Quality

Feature	n-weighted M_1	n-weighted M_2	k[a]	n[b]
Overall	.427	.532	74	1,170
Excellent	.609	.701	56	552
Good	.264	.333	18	618

ANOVA results

Source	df	F	p	η^{2} [c]
Study quality	1, 72	8.01	< .05	.4820

[a] k is the number of correlations.
[b] n is the number of subjects.
[c] η^2 is eta squared, which is defined as: (sum of squares $_{\text{main effect}}$)/(sum of squares $_{\text{total}}$).

17), swimming is statistically significantly different from those of both the bicycle ergometer and treadmill (Table 19), maximal exertion is statistically significantly different from the other four protocols (Table 20), and excellent studies are statistically significantly different from good ones (Table 21).

Pairwise Interactions

Heart Rate

All 15 possible pairwise two-way ANOVAs were calculated. Eight out of the 15 interactions are statistically significant (Table 22): gender × RPE scale, gender × exercise type, gender × exercise protocol, RPE scale × exercise type, RPE scale × study quality, exercise type × exercise protocol, exercise type × study quality, and exercise protocol × study quality (for descriptive statistics of the statistically significant interactions, see Table A.7). To take gender × RPE scale as an example, not all types of RPE scales used are equally effective for males and females when using the weighted validity coefficients as the dependent variable in correlating heart rate with ratings of perceived exertion. Similarly, not all exercise types are equally effective for all types of protocols used when using the weighted validity coefficients as the dependent variable in correlating heart rate with ratings of perceived exertion. Hence, given that over half of the pairwise interactions (8 out of 15) reveal statistically significant interactions, and that "the absence of interactions is the statistical justification for generalizability" (italics deleted, Glass & Hopkins, 1996, p. 485), this analysis corroborates the lack of homogeneous correlations as revealed in the preceding analysis.

Table 22

Statistically Significant Two-Way Interactions for Heart Rate and Oxygen Uptake

Source	df	F[a]	η^2
Heart rate			
Gender × RPE Scale	2, 136	3.44	.0468
Gender × type	3, 134	6.04	.1042
Gender × protocol	3, 134	10.52	.1726
RPE Scale × type	2, 151	5.25	.0591
RPE Scale × quality	2, 144	6.01	.0621
Type × protocol	6, 146	3.61	.1139
Type × quality	4, 141	2.46	.0499
Protocol × quality	4, 141	10.48	.1814
Oxygen uptake			
Gender × quality	1, 59	8.11	.1156
Type × protocol	4, 63	8.85	.3089

[a] All interactions are statistically significant at $p < .05$.

Note that with the exception of fitness level, all other study characteristics are involved in at least one of the eight statistically significant interaction terms. Such lack of generalizability, as revealed by the prevalence of statistically significant interactions, precludes the identification of moderator variables--variables identified as being responsible for the observed heterogeneity. Because all four fitness level features were found to be statistically significant in the χ^2 test of homogeneity, none of the fitness levels can be considered as moderator variables either.

Blood Lactate

All possible 15 two-way ANOVAs were calculated between each of two study characteristics to assess the statistical significance of their respective effects on the weighted correlation ($n \times r$) and blood lactate.

Although Table 22 does not reveal any statistically significant interactions, it still cannot be said that the validity coefficients are generalizable across studies. Three two-way ANOVAs (gender × RPE scale, fitness level × exercise protocol, and RPE scale × exercise protocol) were calculated with an inadequate number of df in the model. No interaction term could be generated. Based on the preceding homogeneity test and the inconclusive results of these three two-way ANOVAs, the validity coefficients between ratings of perceived exertion and blood lactate cannot be generalized across studies. Nor can any moderator variable be identified among the study features.

Oxygen Uptake

All possible 15 two-way ANOVAs were calculated between each of two study characteristics to assess the statistical significance of their respective effects on the weighted correlation ($\underline{n} \times \underline{r}$) and oxygen uptake.

Only 2 out of the possible 15 two-way ANOVAs reveal statistically significant interaction terms: gender × study quality and exercise type × exercise protocol (Table 22; for descriptive statistics of the statistically significant interactions, see Table A.8). However, four other two-way ANOVAs are inconclusive in the interaction terms due to an inadequate number of $\underline{df}$ specified in the model: gender × RPE scale, RPE scale × exercise type, RPE scale × exercise protocol, and RPE scale × study quality. As was seen above, for oxygen uptake, only two types of RPE scales were used (Table 18): the 15-point and the category ratio scales. Therefore, in light of the ubiquitous heterogeneity, the two statistically significant interaction terms and the inconclusive results of the aforementioned two-way ANOVAs, the mean validity coefficients cannot be generalized across studies. Again, nor can any moderator variables be identified among the study features.

Correction for Artifacts

The two artifacts (Table 2) were not combined into a single product; but rather, they were calculated individually as a mean for each study feature: $\sigma_p^2 = \sigma_r^2 - \sigma_e^2$, which, recall from Table 2, is equation (8) minus artifact no. 1, and an attenuation factor that corrects for bias, $\underline{r}_i = \underline{r}_o/[(2N-2)/(2N-1)]$ (artifact no. 9). Table 23 below lists each of these individual $\underline{n}$-weighted artifact-corrected correlations for each study feature for heart rate, blood lactate, and for oxygen uptake, respectively. As can be

Table 23

Artifact-Corrected Validity Coefficients of Each Study Feature for the Three Criterion Variables

	Heart rate		Blood lactate		Oxygen uptake	
Feature	M_1	M_{CB}[a]	M_1	M_{CB}	M_1	M_{CB}
Heart Rate Overall	.532	.607	----[b]	----	----	----
Blood Lactate Overall	----	----	.530	.614	----	----
Oxygen Uptake Overall	----	----	----	----	.427	.649
Gender Overall	.531	.607	.530	.614	.425	.649
Males	.525	.629	.463	.520	.577	.697
Females	.538	.570	.778	.782	.123	.598
Fitness Level Overall	.545	.607	.530	.614	.376	.649
Sedentary	.379	.396	----	----	----	----
Healthy-inactive	.521	.620	.674	.745	.509	.560
Active	.593	.692	.499	.581	.241	.645
Highly fit	.660	.665	.610	.622	.582	.644
RPE Scale Overall	.532	.607	.530	.614	.427	.649
15-point	.535	.616	.721	.731	.408	.652
21-point	.605	.642	----	----	----	----
9-point	.549	.646	.457	.458	----	----
Category ratio	.450	.447	.412	.425	.549	.600

(table continues)

Feature	Heart rate		Blood lactate		Oxygen uptake	
	M_1	M_{CB}[a]	M_1	M_{CB}	M_1	M_{CB}
Exercise Type Overall	.533	.607	.530	.614	.427	.649
Bicycle ergometer	.577	.607	.477	.551	.432	.706
Treadmill	.482	.574	.712	.740	.331	.403
Track running	.332	.513	----	----	----	----
Swimming	.778	.774	.755	.775	.858	.866
Protocol Overall	.533	.607	.530	.614	.427	.649
Progressive continuous	.583	.615	.494	.601	.413	.656
Progressive intermittent	.698	.781	.735	.730	.565	.626
Random intermittent	.450	.354	.457	.458	.750	.792
Maximal exertion	.841	.933	.480	.486	-.055	.504
Submaximal exertion	.457	.568	.582	.580	.532	.675
Study Quality Overall	.528	.607	.530	.614	.427	.649
Excellent	.530	.619	.523	.617	.609	.684
Good	.587	.587	.600	.601	.264	.541
Poor	.304	.357	----	----	----	----

[a] CB is corrected for bias.

[b] The values are either not applicable or no studies reported any values for these study features.

seen from Table 23, the correlations corrected for bias (artifact no. 9) are higher than their uncorrected predecessors, although by a negligible amount in some cases. In addition, the average sampling error variance (Equation 8 - artifact no. 1) is reduced, albeit by a negligible amount.

Study Effects Meta-Analysis

To assess for possible differences among studies themselves, where the study is treated like any other independent variable, a study effects meta-analysis was conducted. The mean validity coefficients for all three criterion variables, as well as for overall (all 60 studies), are comparable in magnitude (Table 24), ranging from approximately .61 to approximately .66.

Table 24

Mean Validity Coefficients for the Study Effects Meta-Analysis

Source	k[a]	M	SD	Variance
Heart rate	49	.657	.233	.054
Blood lactate	15	.642	.210	.044
Oxygen uptake	23	.609	.359	.129
Overall	60	.627	.277	.077

[a] k is the number of studies.
For all three criterion variables, the means are provided in Table A.9 by study.

CHAPTER VI

DISCUSSION AND CONCLUSIONS

Test of Homogeneity

General

A test of homogeneity was applied to all three criterion variables to determine if the validity coefficients are more heterogeneous than would be expected on the basis of sampling variability alone (Table 3). If they turned out to be homogeneous, then the meta-analysis would be finished and there would be no need to continue. All three criterion variables, however, revealed that the validity coefficients are more heterogeneous than sampling variability alone would have predicted.

Given the relatively large number of validity coefficients that are combined under each of the three criterion variables and their attendant study features and the sensitivity of the χ^2 statistic to this number (Hedges, 1981, 1982), it is not surprising that the validity coefficients for almost every study feature under each of the three criterion variables are heterogeneous. Even from a perfunctory glance at Tables A.1, A.2, and A.3, one could guess at the extent of the heterogeneity of the validity coefficients: for heart rate (Table A.1), the range of the validity coefficients between heart rate and RPEs ($rs_{RPE, heart\ rate}$) is -.30 (Winborn, Myers, & Mulling, 1988) to .999 (e.g., Hassmen, 1990) and the range of n is 3 (Edwards et al., 1972) to 128 (Lollgen, Ulmer, & Nieding, 1977; $r_{n,r}$ = -.244); for blood lactate (Table A.2), the range of

$r s_{RPE, \text{blood lactate}}$ is .08 (Borg et al., 1985) to .990 (Zeni et al., 1996) and the range of n is 3 (Edwards et al., 1972) to 50 (Robertson et al., 1979; $r_{n,r}$ = -.554); and for oxygen uptake (Table A.3), the range of $r s_{RPE, \text{oxygen uptake}}$ is -.61 (Rudolph & McAuley, 1996) to .999 (Ueda, Kurokawa, Kikkawa, & Choi, 1993) and the range of n is 3 (Edwards et al., 1972) to 128 (Butts, 1982; $r_{n,r}$ = -.520). Overall, the correlation between n and the validity coefficients is -.337. The negative correlations between sample size and validity coefficients are consistent with small samples producing spuriously larger relationships with each respective criterion variable by chance alone. Such effects would be less likely in studies using large sample sizes, since the subject pool will likely be more heterogeneous. Although these validity coefficients and sample sizes of each of these studies represented possible outliers, for the purposes of this meta-analysis, none of them were excluded as they all met the inclusionary criteria and were rated at least "excellent" or "good" in the study characteristic of study quality. As can be seen from Tables A.1, A.2, and A.3, therefore, the wide range of validity coefficients and sample sizes would greatly underscore the heterogeneity observed in this meta-analysis. Even if each study used highly homogeneous subjects, there would probably still be a great deal of between-group (study) heterogeneity. In addition, it is generally accepted that "validity coefficients derived from scores of homogeneous groups will be lower than those from scores of heterogeneous groups" (Worthen, Borg, & White, 1993, p. 191). Thus, because the studies in this meta-analysis employed samples of subjects differing widely in terms of their demographics, ranging from fairly homogeneous groups in some studies to fairly heterogenous groups in other studies, the observed heterogeneity, as well as the values of the mean validity coefficients themselves, are not surprising.

In the earlier studies of Borg and others, however, only very homogeneous subject groups, such as those with resting and exercising blood pressures that fell within a certain interval, specific occupations (e.g., lumber workers), specific age criteria (Borg & Linderholm, 1967, 1970), or lean and/or obese subjects (Skinner et al., 1973b) were used. Borg (1973) originally devised and calibrated the 15-point RPE Scale so that each numerical point × 10 is comparable to heart rate (in beats per minute) for young-to-middle-aged men (Noble & Robertson, 1996). As shown in Table A.1, these earlier studies of Borg and others generated $rs_{\text{RPE, heart rate}}$ within the approximate range of .70 to .90 (e.g., r = .85; Gamberale, 1972).

Others (e.g., Bar-Or et al., 1972; Gamberale, 1972), on the other hand, used fairly heterogeneous samples, differing, say, in terms blood pressure, percent body fat, serum cholesterol, and number of cigarettes smoked per day, that may be good for generalizability to the general population, but not very sound from an experimental perspective; that is, good external validity, but poor internal validity. For heart rate, blood lactate, and oxygen uptake relationships with ratings of perceived exertion, such disparate findings among different studies are probably due to differences in environmental conditions (Skinner et al., 1973b), exercise tasks (Sargeant & Davies, 1973, 1977; Zeni et al., 1996), protocols (e.g., Brown et al., 1996b; Rudolph & McAuley, 1996), age (Borg & Linderholm, 1967), training (Ekblom & Goldbarg, 1971; Winborn et al., 1988), illness (Borg & Linderholm, 1967; Turkulin, Zamlic, & Pegan, 1977), or even the presence or lack of prior experience of the subjects with the Borg RPE Scale (Miller et al., 1985). Additionally, Ulmer, Janz, and Lollgren (1977) showed just as close a relationship between workload and ratings of perceived exertion scores as to heart rate and ratings of perceived exertion scores.

It is not surprising, therefore, that the mean $\underline{r}s_{\text{RPE, heart rate}}$ for the study characteristic of RPE scales is not higher (range, .45 to .60, Table 6), comparable to that found by Borg and his colleagues. Overall, in this analysis, the category-ratio scale is the poorest predictor of the $\underline{r}_{\text{RPE, heart rate}}$, whereas the 21-point scale is the best. However, depending on which n-weighted mean correlation is used (M_1 or M_2), all four scales are comparable. Individually, by study, however, several studies of the Borg Scale(s) have revealed that the 15-point scale is superior to other modifications of it (Arstila, Wendelin, Vuori, & Valimaki, 1974; Borg, 1973; Table A.2). And both heart rate and blood lactate, when used in conjunction with the category-ratio scale, are not terribly good predictors of perceived exertion (Bloem et al., 1991).

Outliers

The presence of outliers deserves some mention. Stem-and-leaf and box plots of each of the three criterion variables reveal that there are two outliers for heart rate ($\underline{r}$ = -.31, $\underline{r}$ = -.30, Winborn et al., 1988, Figure 2), no outliers for blood lactate (Figure 3), and four outliers for oxygen uptake ($\underline{r}$ = -.21, Butts, 1982; $\underline{r}$ = -.39, Edwards et al., 1972; $\underline{r}$ = .08, Kolkhorst et al., 1996; $\underline{r}$ = -.61, Rudolph & McAuley, 1996; Figure 4). However, as mentioned earlier, because each of these five studies was conducted in an

```
Stem Leaf                                  #      Boxplot

  10 0000000000000                        13         |
   9 0126788                               7         |
   8 00000112223334445555667777888889     32      +-----+
   7 0223344445557778                     16      |     |
   6 0111112233333444456789               22      *--+--*
   5 002223344466777888999                21      |     |
   4 1113356667777888                     16      +-----+
   3 02345779                              8         |
   2 334455556789                         12         |
   1 00146788                              8         |
   0 1133                                  4         |
  -0
  -1
  -2
  -3 10                                    2         0
     ----+----+----+----+----+----+--
     Multiply Stem Leaf by .10
```

Figure 2. Stem-and-leaf plot and boxplot showing the distribution of validity coefficients for heart rate. In the boxplot, a circle represents an outlier(s).

```
Stem Leaf              #         Boxplot

9 099                  3            |
8 0149                 4            |
7 12344578             8         +-----+
6 11234789             8         *--+--*
5 34                   2         |     |
4 025557               6         +-----+
3 14                   2            |
2                                   |
1 58                   2            |
0 8                    1            |
----+----+----+----+
Multiply Stem Leaf by .10
```

Figure 3. Stem-and-leaf plot and boxplot showing the absence of outliers for validity coefficients for blood lactate. In the boxplot, a circle represents an outlier(s).

```
Stem Leaf                        #              Boxplot

  10 0                           1                 |
   9 00000012347                11                 |
   8 02233667788999             14              +------+
   7 345556777777889            15              *------*
   6 0011255678888              13              |  +   |
   5 122667                      6              +------+
   4 78                          2                 |
   3 1369                        4                 |
   2 2                           1                 |
   1 245                         3                 |
   0 8                           1                 0
  -0
  -1
  -2 1                           1                 0
  -3 9                           1                 *
  -4
  -5
  -6 1                           1                 *
     ----+----+----+----+
Multiply Stem Leaf by .10
```

Figure 4. Stem-and-leaf plot and boxplot showing the distribution of validity coefficients for oxygen uptake. In the boxplot, a circle represents an outlier(s) and asterisk represents an extreme outlier(s).

exemplary fashion and was rated either "excellent" or "good" (Tables A.1 and A.3), none of these outliers were excluded from the meta-analysis. Although it is recognized that the ratings of "excellent," "good," and "poor" under the feature of study quality was arbitrarily determined a priori by the author of this meta-analysis (Table 1) and subsequently assigned to each individual study, none of these studies reported any methodological, design, or experimental procedures that would have warranted excluding them from the meta-analysis.

Heart Rate

Fitness levels and maximal exertion in which most or all of the major muscle groups are involved seem to be the best predictors of the relationship between ratings of perceived exertion and heart rate. Intuitively, fitness level has been shown to have a minor, contrary dependency on low and high degrees of stress. That is, when the stress range is high, ratings of perceived exertion tend to decrease with increasing fitness levels (Ulmer et al., 1977). Counter-intuitively, however, highly experienced fit male and females had higher $\underline{r}s_{\text{RPE, heart rate}}$ than their low-experienced counterparts at 70% maximal workload; at 50% workload, highly experienced females exhibited no $\underline{r}s_{\text{RPE, heart rate}}$ (.01), while both low experienced genders exhibited moderate $\underline{r}s_{\text{RPE, heart rate}}$; and at 30% workload, the low-experienced subjects (both genders) exhibited negative $\underline{r}s_{\text{RPE, heart rate}}$ (-.30), whereas both highly experienced genders exhibited moderate $\underline{r}s_{\text{RPE, heart rate}}$ (Winborn et al., 1988; Table A.1). These results indicate that differences in ratings of perceived exertion accuracy may be influenced by gender, but that prior exposure to athletic experiences (fitness levels) appears to override most gender differences.

Involvement of the major muscle masses are also potent predictors of $\underline{r}s_{\text{RPE, heart rate}}$. Hassmen (1990) stated:

> Because running generally uses larger muscle masses than does cycling, running usually exhibits higher heart rates than does cycling at the same subjective intensity. . . . The more intense use of smaller muscle masses during cycling results in higher ratings of perceived exertion, even though the heart rates are comparably lower. (p. 450)

Others (Sargeant & Davies, 1973, 1977), however, found that ratings of perceived exertion are closely associated with heart rate (Table A.1) and oxygen uptake (Table A.3), regardless of the amount of muscle mass involved.

Blood Lactate

More than the other two criterion variables, blood lactate is the most difficult to measure from a technical perspective and the most difficult to interpret from a metabolic perspective. Kay and Shepard (1969) found one of the lowest correlations between blood lactate and ratings of perceived exertion ($rs_{RPE, blood lactate}$, .15, Table A.2) in this meta-analysis, perhaps due to the rather questionable way in which they collected the lactate--by heating the subject's finger to 45°C (113°F) before puncture; lactate may be temperature-sensitive and may degrade at 45°C. Also, blood withdrawal from the finger may not be the most accurate, since it is such a distant extremity. A more central location (e.g., the arm) would be better.

In addition, it is well known that blood lactate increases quadratically, rather than in a linear fashion as a function of ratings of perceived exertion (Bergh, Danielsson, Wennberg, & Sjodin, 1986; Borg, Van den Burg et al., 1987; Borg et al., 1985; Demello et al., 1987; Noble et al., 1983; Pedersen & Welch, 1977; Ueda et al., 1993). However, none of the articles examined mentioned any sort of linear transformation of their data, where ratings of perceived exertion are squared and logarithms (base 10) of lactate concentrations are calculated. If some studies transformed their data, while others did not, this could represent a profound contribution to the observed heterogeneity.

Physiologically, central versus peripheral signals of perceived exertion (where subjects are instructed to think about certain parts of their bodies, e.g., the torso [central], arms, legs [peripheral] when performing a certain exercise task; e.g., Robertson et al., 1979; Skrinar et al., 1983) may account, to some degree, for the lower-than expected (mean) $\underline{r}s_{\text{RPE, heart rate}}$, because most investigators do not instruct their subjects to specifically pay attention to and feel where on their bodies, peripherally or centrally, exertion, stress, strain, pain, or fatigue is occurring. For example, in swimming, when such differentiated ratings of perceived exertion (legs, arms, cardiac frequency ratings, or respiratory rate) were used (Ueda et al., 1993), various physiological measures were found to correlate moderately to highly with certain differentiated ratings of perceived exertion. It should be noted, that, in turn, each of these differentiated ratings of perceived exertion correlated very highly ($\underline{r}$ = .97, .98) with overall (global) ratings of perceived exertion (Ueda et al., 1993). Mihivec (1981) has suggested that blood lactate influences ratings of perceived exertion primarily at higher exercise intensities, such as when pain is felt, but makes only a minimal contribution at lower intensities. To the extent that higher exercise intensities mean recruitment of larger muscle masses, it makes sense that in this meta-analysis, swimming (Kurokawa & Ueda, 1992) and treadmill use (Hassmen, 1990, Table 13) had among the highest $\underline{r}s_{\text{RPE, blood lactate}}$. Indeed, pain and discomfort (fatigue) cannot be separated from other physical estimates (Pandolf, Cafarelli, Noble, & Metz, 1972; Toner et al., 1986), which, in turn, cannot be separated from perception of effort (Noble & Robertson, 1996; Ulmer et al., 1977). Consistent with the mean

$\underline{r}s_{\text{RPE, blood lactate}}$ found in this meta-analysis, the validity coefficient between ratings of perceived exertion and discomfort has been reported to be .59 (Kamon, Pandolf, & Cafarelli, 1974).

Oxygen Uptake

As mentioned above, the heterogeneity, as well as overall lower range of mean correlation between oxygen uptake and ratings of perceived exertion ($\underline{r}s_{\text{RPE, oxygen uptake}}$), can best be explained by the fact that four physiologically different oxygen uptake measures were combined into one global measure. The rationale for this is that from a meta-analytic perspective, subtly different measures do not always translate into substantively and practically different measures in the real world (Glass et al., 1981). Indeed, recall that the units of each of these four measures of oxygen uptake are interconvertible. From a measurement point of view, subjects probably feel their breathing (which includes oxygen uptake) more than they feel their heart rate and certainly more than they feel their blood lactate (unless pain is involved). Support for this comes from earlier investigations using cycling (e.g., Cafarelli & Noble, 1976; Michael & Eckardt, 1972; Pandolf & Noble, 1973; Stamford & Noble, 1974) or treadmill (Demello et al., 1987) when oxygen uptake, per se, is not the major influence in the subjective rating of work stress.

The heterogeneity of $\underline{r}s_{\text{RPE, oxygen uptake}}$ revealed in this meta-analysis can be further traced to the wide spectrum of "percent variance accounted for" in several of the studies examined. For example, during moderate to heavy work, ratings of perceived exertion are significantly affected by oxygen consumption (Allen & Pandolf, 1977), such as by $\%VO_{2\,max}$ which has been found to explain nearly 80% (Brown et al., 1996b), 69% (Rudolph & McAuley, 1996), and 30% (Butts, 1982) of the variance in perceived exertion. Such disparate findings lend support to the possibility that ratings of perceived exertion do not depend on $\%VO_{2\,max}$ (Noble & Robertson, 1996, p. 113).

Conclusions

The ability of subjects to feel oxygen uptake (or, more generally, breathing) more keenly than they feel the other two criterion variables makes the former more susceptible to the confounds of psychological factors (Borg, Skinner, & Bar-Or, 1972; Clapp & Little, 1994; Hardy & Rejeski, 1989; Hassmen, 1990; Miller et al., 1985; Morgan, 1973; Morgan & Pollack, 1977; Rudolph & McAuley, 1996; Williams & Eston, 1986; Winborn et al., 1988). It is extremely difficult to establish any sort of pattern when such factors are involved. For example, highly active subjects have been shown to have reduced state anxiety after exercising, although more inactive subjects did not (Dishman et al., 1994). Curiously, these investigators also found that both highly active subjects and inactive ones did not differ in their ratings of perceived exertion, despite their relative differences in exercise intensity (Dishman et al., 1994).

The results of this meta-analysis indicate that no one physiological criterion variable (heart rate, blood lactate, or oxygen uptake) is that much better in its relationship with ratings of perceived exertion than any other. Several study features, however, have been shown to be notably superior, such as the use of highly fit subjects, or highly stressful (maximal exertion) or unusual (swimming) tasks. These features, when the subject is noticably challenged, are potent predictors of the ratings of perceived exertion-criterion variable relationship. Thus, if forced to provide a choice, the 15-point Borg RPE Scale would be best used with a group of highly fit young-to-middle-aged men, performing tasks that are either relatively unusual (e.g., swimming) or require extreme exertion. Under such conditions, the relationship between both heart rate and blood lactate with ratings of perceived exertion will be at least moderately high.

Furthermore, it is usually better to have as many measures as possible. In this meta-analysis, because there is not enough documentation on any others (e.g., hormones, creatine) to warrant a meta-analysis, only three criterion variables were examined. However, if it is possible to use at least the three criterion variables examined in this meta-analysis in future studies, then it would well behoove the investigator to do so. Combined, heart rate is a better predictor of ratings of perceived exertion and blood lactate is a better predictor of perceived exertion-pain than is each taken alone (Borg et al., 1985; Ljunggren, Ceci, & Karlsson, 1987; Steed et al., 1994). Because physiologically, heart rate, lactate, and oxygen are all well-known to be central mediators of the complex biochemical and physiological processes of metabolism, it is ideal to measure at least all three, rather than each alone.

Mean Validity Coefficients

Heart Rate

The mean n-weighted validity coefficients of the study features for heart rate with RPE scores cover a wide range, for the most part, approximately .30 to .99. Most notable within this range are the study features of highly fit (.815, Table 5), swimming (.834, Table 7), maximal exertion (.985, Table 8), and poorly conducted studies (.30, Table 9). Note that high mean validity coefficients are obtained when either the task is relatively unusual (e.g., swimming), when the entire body is involved (Kurokawa & Ueda, 1992), the subjects themselves are extreme in some way (highly fit), or the task requires the subjects to maximally exert or strain themselves (maximal exertion). Note also that poorly conducted studies find low correlations between RPE scores and heart rate.

Consistently, the four study characteristics to which each of these four features belong, respectively, also have statistically significant F values and relatively large effect sizes (η^2). With the exception of quality, where η^2 is .0433, the effect sizes of statistically significant F values (fitness level, exercise type, and exercise protocol) are an order of magnitude larger than those that did not reach statistical significance (i.e., gender and RPE Scale).

Some of the results are intuitively appealing. For example, under the heading of exercise protocol, maximal exertion is statistically significantly different from all other protocols. On the other hand, some of the results from the Tukey's test are counterintuitive. For example, under the fitness level characteristic (Table 5), contrary to the findings, one would not necessarily expect sedentary individuals to differ from healthy-inactive subjects (see Hassmen, 1990; Jackson, Dishman, LaCroix, Patton, & Weinberg, 1981; Ulmer et al., 1977). Such a counterintuitive result might be attributed to the fact that some studies reported as their method of determining subjects' fitness level simply the subjects' testimony of how fit they consider themselves to be (Gamberale, 1972) or an account of their daily lifestyle (e.g., Turkulin et al., 1977). It is possible that different investigators and subjects harbor different definitions or criteria of what constitutes a sedentary lifestyle, as opposed to, say, a healthy-inactive one. Thus, it might be more appropriate to determine fitness level on the basis of the subjects' percent maximal oxygen uptake values (e.g., Travlos & Marisi, 1996; Winborn et al., 1988).

Blood Lactate

The mean n-weighted validity coefficients of the study features for blood lactate fall approximately within the .40 to .84 range. With this criterion variable, however, there are approximately six study features displaying fairly large mean n-weighted validity coefficients in the .75 to .80 range: females (Table 10), healthy-inactive individuals (Table 11), 15-point Borg RPE Scale (Table 12), treadmill use (Table 13), swimming (Table 13), and the progressive intermittent protocol (Table 14).

Only the study characteristics of gender (Table 10, males vs. females) and RPE Scale (Table 12, 15-point vs. 9-point and category-ratio) display between-group (features) statistical significance and corroboratively larger effect sizes, .316 and .430, respectively. All the other effect sizes of the other study characteristics are much smaller and are not statistically significant.

The superiority of females over males (Table 10) is consistent with findings presented in the literature. For example, Kurokawa and Ueda (1992; Ueda et al., 1993) found that because of their generally greater buoyancy (due to a higher percentage of body fat, as compared to males), women are better able to maximally exert themselves to the point of lactic acid accumulation and its attendant pain and, therefore, provide a stronger relationship between ratings of perceived exertion and blood lactate. Males, on the other hand, generally have higher basal metabolic rates and, therefore, do not accumulate blood lactate as quickly.

Historically, in the early 1970s, the 15-point Borg RPE Scale was a refinement over the original 21-point Borg RPE Scale (Noble & Robertson, 1996). Since then, the 15-point scale has become the most heavily used RPE scale (Tables 6, 12, and 18). Its widespread use, however, does not necessarily mean that it is the most appropriate when blood lactate is the criterion variable. Recall that earlier, it was mentioned that certain physiological variables (e.g., blood lactate) are related to exercise intensity according to nonlinear power functions (Borg, Van den Burg et al., 1987). For example, when lactate is plotted as a function of the Borg RPE Scale (ratings 6-20), the lactate concentration increases about three times more per scale unit at the top of the scale (ratings 16-17, see Figure 1) than at the bottom (see Noble et al., 1983). A different scale, therefore, was needed that would identify fatigue associated with nonlinear physiological responses, such as lactate accumulation.

Therefore, Borg developed a new psychophysical scale so that perception ratings would increase as a positively accelerating function and was termed the category-ratio scale. The nonlinear nature of blood lactate accumulation as a function of power output and perceived exertion has been demonstrated where blood lactate was found to be most appropriately described as a quadratic function (Noble et al., 1983). Perhaps at higher intensities when the glycogenolytic contribution to energy production is highest, there would be little relationship between ratings of perceived exertion and lactate; that is, lactate rises rapidly while ratings of perceived exertion continue to increase linearly with power output.

Oxygen Uptake

The range of mean n-weighted validity coefficients for oxygen uptake, unlike those for heart rate and blood lactate, is a little wider and has more lower validity coefficients, making the general range to be approximately - .055 to .858. Most notable are females (ca. .20, Table 16), active fitness level (ca. .20 to .40, Table 17), and study quality of "good" (ca. .26 to .33, Table 21). There is no substantive correlation between ratings of perceived exertion and one study feature, maximal exertion (Table 20), which is somewhat surprising, given that maximal exertion exhibited very high mean validity coefficients between ratings of perceived exertion and heart rate (recall above, Table 8). Only one study feature, swimming, which engages the entire body and that requires that breathing be coordinated with the placement of the head in and out of water, exhibited a very high mean validity coefficient (ca. .90) between oxygen uptake and ratings of perceived exertion (Table 19). And the study quality of "excellent" displayed moderately high mean validity coefficients (ca. .60 to .70, Table 21).

There are at least two important reasons--one from a measurement perspective, one from a physiological perspective--why the range of mean n-weighted correlations for oxygen uptake is lower than that for heart rate and blood lactate: (a) measurement--oxygen uptake was treated as one global variable, combining four different measures of oxygen uptake (% maximal oxygen uptake, oxygen uptake, respiratory rate, and minute ventilation) into one (recall Methods); and (b) physiological--a lower (mean) oxygen uptake is the prerequisite for the metabolic switch to a primarily anaerobic mode of metabolism (i.e., higher lactate accumulation), and could explain the higher mean correlation between ratings of perceived exertion and blood lactate. The first reason, therefore, might have served to obscure the true relationship between any one of these oxygen uptake measures and ratings of perceived exertion. And, in principle, because all four oxygen uptake measures, either individually or collectively, are correlated with heart rate (Ainsworth et al., 1997; Bloem et al., 1991; Grant et al., 1993; Ueda & Kurokawa, 1991), which, in turn, is more highly correlated with ratings of perceived exertion, then oxygen uptake measures should be correlated with ratings of perceived exertion as well.

Only the study feature of RPE Scale is not statistically significant (Table 18); its effect size ($\eta^2 = .002$) is two orders of magnitude smaller than that for gender, fitness level, exercise type, exercise protocol, and study quality (Tables 16, 17, 19, 20, and 21, respectively).

Pairwise Interactions of Heart Rate

Of the eight statistically significant interaction terms, the most prevalently involved is exercise type, which significantly interacts with four of them (Table 22): gender, RPE scale, exercise protocol, and study quality. As indicated above, swimming is a profound predictor of $\underline{r}s_{\text{RPE, heart rate}}$ and is probably so different under a wide variety of circumstances that it is not consistent across genders, types of Borg RPE Scales, protocols, or quality of study (e.g., Kurokawa & Ueda, 1992). Gender, RPE Scale, exercise protocol, and study quality are involved in three interactions each; and it is surprising that only fitness level is not involved in any interaction, given that this study characteristic was statistically significant in the within-group (characteristic) one-way ANOVA, revealing the potency of highly fit subjects to predict $\underline{r}s_{\text{RPE, heart rate}}$.

Pairwise Interactions of Oxygen Uptake

Only two statistically significant pair-wise interactions were found: that between gender and study quality, and between exercise type and protocol (Table 4). From a physiological perspective, such findings can be traced back to the statistically significant correlations found between blood lactate and ratings of perceived exertion. Recall that this relationship revealed that females (gender), treadmill, swimming (exercise types), and the progressive intermittent protocol yielded sizable mean sample size-weighted validity coefficients. Across gender, exercise type, and exercise protocol, therefore, it makes sense that features subsumed under each of these study characteristics would be involved in some significant interaction in that a switch to a primarily anaerobic mode of metabolism is required (recall above).

Conclusions

Consistent with the results of the test of homogeneity, and based on the mean validity coefficients, no one of the three criterion variables is that much better than any of the other two in predicting ratings of perceived exertion. Overall mean sample size-weighted validity coefficients for heart rate and blood lactate are approximately within the .500 to .600 range, whereas those for oxygen uptake are a little, but not substantively, lower (within the .400 to .500 range). Most notable, however, are those study features that exhibit a high correlation between ratings of perceived exertion and the criterion variable. Such features are highly fit individuals, maximally exerting themselves, in an exercise/work task that is fairly unusual (i.e., swimming).

Correction for Artifacts

As it may be recalled from previous chapters, the tendency for studies from the Borg RPE literature to report quantifiable artifact information is of a sporadic nature only. Because not all studies provide artifact information, therefore, artifact distributions for each of the three criterion variables were calculated. For all three criterion variables, as provided in Table 23, for heart rate, blood lactate, and oxygen uptake, respectively, the artifact corrections for bias are now higher than their uncorrected counterparts. Note that although the bias-corrected correlations are higher than their uncorrected counterparts, the former are not higher than those mean validity coefficients that have been converted to their Fisher's zs, averaged, and then back-transformed to their respective correlations (M_2, Tables 4-21). This is because the correction for bias represents only a statistical bias in the sample correlation as an estimate of the population correlation and is independent of the latter (Hunter & Schmidt, 1990). Fisher's z transformation, on the other hand, tends to give disproportionately more weight to larger values of r than to moderate values (Hunter et al., 1982); and even when back-transformed, the validity coefficients end up being larger--even larger than the bias-corrected validity coefficients.

Because this meta-analysis is using artifact distributions to correct for artifacts, rather than correcting each correlation by study (because not all studies provided quantifiable artifact information), it is neither necessary nor sound to trace the reason down to one study. That is, when validity coefficients are corrected individually, the mean and variance of each corrected validity coefficient are computed. Each study, therefore, could be treated separately from the rest. For an artifact distribution, however, the mean and the variances of the uncorrected coefficients are used as multipliers (in the case of correcting for bias) to eliminate the effects of the bias (Hunter et al., 1982).

In conclusion, then, simply by dividing each mean validity coefficient by a common correction factor (constant), the corrections for bias resulted in higher validity coefficients and lower sampling variances for all study features within each of the three criterion variables.

Study Effects Meta-Analysis

The purpose of a study-effects meta-analysis is to determine if the studies themselves, as the unit of analysis, contribute to the observed results. The only real difference between this particular type of analysis and the Homogeneity Test performed earlier is that in the latter, all weighted validity coefficients from each study were incorporated in the meta-analysis. In the study effects meta-analysis, however, each study contributes only one validity coefficient, whether it is a single coefficient or the mean of several validity coefficients.

Several studies provided validity coefficients that are unusually high (Borg, Van den Burg, et al., 1987; Duncan, Mahon, Gay, & Sherwood, 1996; Hassmen, 1990; Zeni et al., 1996), or unusually low (Butts, 1982; Rudolph & McAuley, 1996; Smith, 1994; Winborn et al., 1988). All four studies reporting validity coefficients of about .99 were included in the study-effects meta-analysis because they were all conducted in an exemplary fashion; that is, they were all rated "excellent" (Bangert-Drowns, 1986; Table A.1). Given that the bulk of the validity coefficients fall within the range of about .30 to .80, it is quite possible that the investigators of these four studies reported correlations based on group means, rather than those based on individual scores. Such mean or "ecological" correlations (Glass & Hopkins, 1996) will usually fall within the range of .90 to .99. Although it is uncertain exactly how these coefficients were calculated, because these studies were rated "excellent" they were included in the meta-analysis. Their inclusion in the study-effects meta-analysis would account for the slightly higher mean validity coefficients for heart rate (.657), as compared to the mean sample size-weighted validity coefficient (.532, .616) reported earlier.

The atypically low validity coefficient reported by Rudolph and McAuley (1996) can be explained from a methodological point of view. These investigators found a moderately large negative validity coefficient between % maximal oxygen uptake and ratings of perceived exertion (r = -.61). Yet, they met all the inclusionary criteria specified in the Methods chapter in this meta-analysis and received a quality mark of "good." Rudolph and McAuley's study examined the correlations among pre- and postexercise self-efficacy, ratings of perceived exertion, and % maximal oxygen uptake. They also found an inverse relationship between subjects' postexercise self-efficacy and their perceptions of effort, suggesting "that higher exercise-related efficacy cognitions are associated with lower perceptions of effort and fatigue during exercise" (p. 220). And although subjects in this study displayed a positive correlation between % maximal oxygen uptake and pre- (r = .59) and postexercise (r = .51) self-efficacy, it is possible that subjects were unconsciously influenced to respond to the Borg RPE Scale as they did the self-efficacy scale. It is also possible that the rather long exercise time of 30 minutes to exhaustion at 60% maximal oxygen uptake contributed to the negative relationship. When the exercise trial was nearly finished (at 29 minutes), subjects provided their ratings of perceived exertion when each breath is deepest (due to exhaustion), but the proportion of oxygen intake was lowest (Noble & Robertson, 1996, p. 112), perhaps due to the increased amount of carbon dioxide in the room (whose dimensions were not specified, or if the task took place outside).

In conclusion, by averaging validity coefficients from each study, resulting in each study contributing only one coefficient, each criterion variable produced a mean validity coefficient that is only slightly higher, but probably not substantively different, from those derived from the Mean Validity Coefficient section reported earlier.

General Conclusions and a Recommendation for Future Directions

The highest n-weighted correlations between ratings of perceived exertion and heart rate are those when the subjects are at the highest end of the fitness continuum, require the subjects to maximally exert themselves, or require fairly unusual circumstances (e.g., swimming). Although the same cannot be said of blood lactate and oxygen uptake, both of these measures also correlated very highly with ratings of perceived exertion when subjects were required to swim.

Virtually all studies examined in this meta-analysis employed fairly homogeneous groups of subjects, although there is still a great deal of between-group (features) differences among them. Additionally, it is generally accepted that the use of more homogeneous groups will produce lower validity coefficients than that of more heterogeneous groups. In light of the negative relationship found between sample size and validity coefficients, it would behoove future investigators to use larger sample sizes. Although doing so might sacrifice some homogeneity within the group of subjects, large samples would not be subject to the spuriously large relationships between ratings of perceived exertion and the criterion variable that smaller samples would more likely engender; and higher true validity coefficients would likely result. That is, maximize subject homogeneity in terms of subject demographics so that when large samples are employed, a safeguard against the imminent decrease in validity coefficients is in place.

When grouped by study characteristics and their attendant features, the n-weighted validity coefficients between ratings of perceived exertion and each of the three criterion variables (heart rate, blood lactate, and oxygen uptake) are heterogeneous. These n-weighted validity coefficients are comparable, albeit a little lower, than when these same validity coefficients are grouped by the study where they originated with each study contributing only one such validity coefficient.

Strangely scant in the perceived exertion literature are studies examining the relationships among several physiological variables, ratings of perceived exertion, and personality factors (although a few were encountered and used in this meta-analysis, e.g., Hardy & Rejeski, 1989; Rudolph & McAuley, 1996). More of these types of studies need to be done to achieve much more homogeneity within subject groups, rather than just relying on surface features, such as fitness level and gender.

The three criterion variables explored in this meta-analysis are some of the major physiological processes that are thought to function as indicators for the respiratory-metabolic signal of exertion. From the perspective of applied research and clinical settings, the application of perceived exertion would work best through a knowledge of the underlying physiological and psychological processes, which the individual subjectively monitors and evaluates. When an individual exercises or performs some other physical labor, this exercise stimulus leads to physiological responses that mediate the intensity of perceptual signals of exertion by acting either individually or collectively in altering tension in skeletal muscle. These changes in muscle tension are monitored through a final common neurophysiological pathway that transmits exertional signals from the motor to the sensory cortex, which consciously interprets this neurophysiological signal as effort sensation. Inherent in this sensory continuum is a feedback loop wherein graded perceptual reports are interactively linked with physiological and neurological events so that the appropriate exertional response under a variety of performance and clinical conditions occurs.

Given the results of this meta-analysis for each of the three criterion variables overall, it appears that no one criterion variable is substantively superior over the other two. As mentioned in Chapter I of this study, the evidence supporting heart rate as a predictor of perceived exertion is not consistent. Although heart rate is easy enough to measure, whether at rest or during exercise, and although cognitively, an individual could guess that his/her heart rate is increased during exercise, it is unlikely that the exact amount of increased beats per minute could be accurately perceived during exercise (or during rest). That is, the psychological mediator in the form of the knowledge that heart rate should increase during exercise might contribute to an increased subjective awareness that perceived exertion should likewise increase.

Although there is ample evidence in this meta-analysis supporting the idea that blood lactate mediates the intensity of exertional perceptions, there is also a fair amount of evidence to the contrary. Nevertheless, like heart rate and three of the four measures of oxygen uptake (see below), a change in blood lactate is also extremely difficult, if not impossible to perceive accurately. Only in the presence of the acute pain normally associated with its accumulation during exercise does the perception of lactic acid buildup become noticeable. But again, the individual cannot measure this increase, but is only aware of it, again perhaps through the psychological mediator of what she/he knows can happen.

As mentioned in the Methods chapter, four of the most highly used measures of respiration were combined under one generic criterion variable of oxygen uptake. Like heart rate and blood lactate (above), whether an individual is at rest or is exercising, it is unlikely that she/he would be able to determine the amount of oxygen inspired with each breath. Furthermore, the same individual performing the same exercise across different environmental settings (e.g., exercise laboratory, outdoors) will most likely be inspiring different amounts of oxygen, depending on the available atmospheric content and the relative distribution of respiratory gases; indeed, the aerobic demand would be higher indoors than it would be outdoors. Therefore, of the four measures of oxygen uptake included in this meta-analysis, it is likely that only respiratory rate (breaths min^{-1}) can serve as the only reliable and valid measure of perceived exertion in both a research and clinical setting. Breaths can easily be counted by both an observer, and the one doing the exercise, and are, therefore, intuitively, probably the best indicator of perceived exertion. Future empirical studies aimed at addressing these questions need to be conducted.

REFERENCES

References marked with an asterisk indicate studies included in the meta-analysis.

Abadie, B. R. (1996). Effect of viewing the RPE scale on the ability to make ratings of perceived exertion. Perceptual and Motor Skills, 83, 317-318.

Adcock, C. J. (1960). A note on combining probabilities. Psychometrika, 25, 303-305.

Ainsworth, B. E., McMurray, R. G., & Veazey, S. K. (1997). Prediction of peak oxygen uptake from submaximal exercise tests in older men and women. Journal of Aging and Physical Activity, 5, 27-38.

*Allen, P. D., & Pandolf, K. B. (1977). Perceived exertion associated with breathing hyperoxic mixtures during submaximal work. Medicine and Science in Sports, 9, 122-127.

Allen, W. K., Seals, D. R., Hurley, B. F., Ehsani, A. A., & Hagberg, J. M. (1985). Lactate threshold and distance-running performance in young and older endurance athletes. Journal of Applied Physiology, 58, 1281-1284.

American College of Sports Medicine. (1991). Guidelines for exercise training and prescription. Philadelphia: Lea & Febiger.

Aminoff, T., Smolander, J., Korhonen, O., & Louhevaara, V. (1996). Physical work capacity in dynamic exercise with differing muscle masses in healthy young and older men. European Journal of Applied Physiology and Occupational Physiology, 73, 180-185.

*Anshel, S. L., Anshel, M. H., & Gregory, W. L. (1991). The effect of type A and type B fitness level on ratings of perceived exertion and physical fatigue by females. The Australian Journal of Science and Medicine in Sport, 23, 106-110.

*Arstila, M., Wendelin, H., Vuori, I., & Valimaki, I. (1974). Comparison of two rating scales in the estimation of perceived exertion in a pulse-conducted exercise test. Ergonomics, 17, 577-584.

*Bacharach, D. W. (1984). Influence of cardiovascular condition of women on perceived exertion during submaximal workloads. Unpublished manuscript, Texas A & M University, College Station.

Bangert-Drowns, R. L. (1986). Review of developments in meta-analytic methods. Psychological Bulletin, 99, 388-399.

*Bar-Or, O., Skinner, J. S., Buskirk, E. R., & Borg, G. (1972). Physiological and perceptual indicators of physical stress in 41- to 60-year-old men who vary in conditioning level and in body fatness. Medicine and Science in Sports, 4, 96-100.

Bergh, U., Danielsson, U., Wennberg, L., & Sjodin, B. (1986). Blood lactate and perceived exertion during heat stress. Acta Physiologica Scandinavica, 126, 617-618.

Billat, L. V. (1996). Use of blood lactate measurements for prediction of exercise performance and for control of training. Recommendations for long-distance running. Sports Medicine, 22, 157-175.

Blair, S. N., Kohl, H. W. K., III, Paffenbarger, R. S., Jr., Clark, D. G., Cooper, K. H., & Gibbons, L. W. (1989). Physical fitness and all-cause mortality: A prospective study of healthy men and women. Journal of the American Medical Association, 262, 2395-2401.

*Bloem, P. J. N., Goessens, L. M. G. C., Zamparo, P., Sacher, M., Paviotti, R., & di Prampero, P. E. (1991). Effects of order of presentation of exercise intensities and of sauna baths on perceived exertion during treadmill running. European Journal of Applied Physiology and Occupational Physiology, 62, 204-210.

Borg, G. A. V. (1962). Physical performance and perceived exertion. Lund, Sweden: Gleerup.

*Borg, G. A. V. (1973). Perceived exertion: A note on "history" and methods. Medicine and Science in Sports, 5, 90-93.

Borg, G. A. (1982). Psychophysical basis of perceived exertion. Medicine and Science in Sports and Exercise, 14, 371-381.

Borg, G., & Linderholm, H. (1967). Perceived exertion and pulse rate during graded exercise in various age groups. Acta Medica Scandinavica Supplementum, 472, 194-206.

Borg, G., & Linderholm, H. (1970). Exercise performance and perceived exertion in patients with coronary insufficiency, arterial hypertension and vasoregulatory asthenia. Acta Medica Scandinavica, 187, 17-26.

Borg, G., Hassmen, P., & Lagerström, M. (1987). Perceived exertion related to heart rate and blood lactate during arm and leg exercise. European Journal of Applied Physiology and Occupational Physiology, 65, 679-685.

*Borg, G., Karlsson, J-G., & Ekelund, L-G. (1977). A comparison between two work tests controlled subjectively and by heart rate. In G. Borg (Ed.), Physical work and effort (pp. 239-354). Oxford: Pergamon Press.

*Borg, G., Ljunggren, G., & Ceci, R. (1985). The increase of perceived exertion, aches and pains in the legs, heart rate and blood lactate during exercise on a bicycle ergometer. European Journal of Applied Physiology and Occupational Physiology, 54, 343-349.

*Borg, G., Skinner, J. S., & Bar-Or, O. (1972). Self-appraisal of physical performance capacity. Baltimore: Williams & Wilkins (ERIC Document Accession No. ED 083-264)

*Borg, G., Van den Burg, M., Hassmen, P., Kaijser, L., & Tanaka, S. (1987). Relationships between perceived exertion, HR and Hla in cycling, running and walking. Scandinavian Journal of Sports Science, 9, 69-77.

Brown, S. P., Chitwood, L. F., Beason, K. R., & McLemore, D. R. (1996a). Perceptual responses to deep water running and treadmill exercise. Perceptual and Motor Skills, 83, 131-139.

*Brown, S. P., Chitwood, L. F., Beason, K. R., & McLemore, D. R. (1996b). Physiological correlates with perceived exertion during deep water running. Perceptual and Motor Skills, 83, 155-162.

Burlew, L. D., Jones, J., & Emerson, P. (1991). Exercise and the elderly: A group counseling process. Journal for Specialists in Group Work, 16, 152-158.

*Butts, N. K. (1982). Physiological profiles of high school females cross country runners. Research Quarterly for Exercise and Sport, 53, 8-14.

Cafarelli, E. (1982). Peripheral contributions to the perception of effort. Medicine and Science in Sports and Exercise, 14, 382-389.

Cafarelli, E., & Noble, B. J. (1976). The effect of inspired carbon dioxide on subjective estimates of exertion during exercise. Ergonomics, 19, 581-589.

Carton, R. L., & Rhodes, E. C. (1985). A critical review of the literature on rating scales for perceived exertion. Sports Medicine, 2, 198-202.

*Clapp, J. F., III., & Little, K. D. (1994). The physiological response of instructors and participants to three aerobics regimens. Medicine and Science in Sports and Exercise, 26, 1041-1046.

*Costa, G., & Gaffuri, E. (1977). Studies of perceived exertion rate on bicycle ergometer in conditions reproducing some aspects of industrial work (shift work; noise). In G. Borg (Ed.), Physical work and effort (pp. 297-305). Oxford: Pergamon Press.

*Demello, J. J., Cureton, K. J., Boineau, R. E., & Singh, M. M. (1987). Ratings of perceived exertion at the lactate threshold in trained and untrained men and women. Medicine and Science in Sports and Exercise, 19, 354-362.

Dishman, R. K. (1994). Prescribing exercise intensity for healthy adults using perceived exertion. Medicine and Science in Sports and Exercise, 26, 1087-1094.

Dishman, R. K., Farquhar, R. P., & Cureton, K. J. (1994). Responses to preferred intensities of exertion in men differing in activity levels. Medicine and Science in Sports and Exercise, 26, 783-790.

Dunbar, C. C., & Bursztyn, D. A. (1996). The slope method for prescribing exercise with ratings of perceived exertion (RPE). Perceptual and Motor Skills, 83, 91-97.

Dunbar, C. C., Glickman-Weiss, E. L., Edwards, W. W., Conley, P., & Quiroz, A. (1996). Three-point method of prescribing exercise with ratings of perceived exertion is valid for cardiac patients. Perceptual and Motor Skills, 83, 384-386.

Dunbar, C. C., Kalinski, M. I., & Robertson, R. J. (1996). A new method for prescribing exercise: Three-point ratings of perceived exertion. Perceptual and Motor Skills, 82, 139-146.

Dunbar, C. C., Robertson, R. J., Baum R., Blandin, M. F., Metz, K., Burdett, R., & Goss, F. L. (1992). The validity of regulating exercise intensity by ratings of perceived exertion. Medicine and Science in Sports and Exercise, 24, 94-99.

*Duncan, G. E., Mahon, A. D., Gay, J. A., & Sherwood, J. J. (1996). Physiological and perceptual responses to graded treadmill and cycle exercise in male children. Pediatric Exercise Science, 8, 251-258.

Dustman, R. E., Emmerson, R. Y., Ruhling, R. O., Shearer, D. E., Steinhaus, L. A., Johnson, S. C., Bonekat, H. W., & Shigeoka, J. W. (1990). Age and fitness effects of EEG, ERPs, visual sensitivity, and cognition. Neurobiology of Aging, 11, 193-200.

*Edwards, R. H. T., Melcher, A., Hesser, C. M., Wigertz, O., & Ekelund, L.-G. (1972). Physiological correlates of perceived exertion in continuous and intermittent exercise with the same average power output. European Journal of Clinical Investigation, 2, 108-114.

Ekblom, B., & Goldbarg, A. N. (1971). The influence of physical training and other factors on the subjective rating of perceived exertion. Acta Physiologica Scandinavica, 83, 399-406.

Emery, C. F., & Blumenthal, J. A. (1991). Effects of physical exercise on psychological and cognitive functioning of older adults. Annals of Behavioral Medicine, 13, 99-107.

Eston, R. G., & Williams, J. G. (1986). Exercise intensity and perceived exertion in adolescent boys. British Journal of Sports Medicine, 20, 27-30.

Evans, B., Hopkins, D., & Toney, T. (1996). Metabolic response to the half-mile AAHPERD functional fitness walk test in older adults. Journal of Aging and Physical Activity, 4, 80-89.

Gage, N. L. (1978). The scientific basis of the art of teaching. New York: Teachers College Press.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). Educational research. New York: Longman.

*Gamberale, F. (1972). Perception of exertion, heart rate, oxygen uptake and blood lacate in different work operations. Ergonomics, 15, 545-554.

Garcin, M., Cravic, J. Y., Vandewalle, H., & Monod, H. (1996). Physiological strains while pushing or hauling. European Journal of Applied Physiology, 72, 478-482.

Gilbert, J., McPeek, B., & Mosteller, F. (1977). Statistics and ethics in surgery and anesthesia. Science, 198, 684-689.

Glass, G. V. (1976). Primary, secondary and meta-analysis of research. Educational Researcher, 5, 3-8.

Glass, G. V. (1977). Integrating findings: The meta-analysis of research. Review of Research in Education, 5, 351-379.

Glass, G. V., & Hopkins, K. D. (1996). Statistical methods in education and psychology. Needham Heights, MA: Allyn & Bacon.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). Meta-analysis in social research. Beverly Hills, CA: Sage.

*Goslin, B. R., & Rorke, S. C. (1986). The perception of exertion during load carriage. Ergonomics, 29, 677-686.

Grant, S., Armstrong, G., Sutherland, R., Wilson, J., Aitchison, T., Paul, E., & Henderson, S. (1993). Physiological and psychological responses to a university fitness session. British Journal of Sports Medicine, 27, 162-166.

*Hardy, C. J., & Rejeski, W. J. (1989). Not what, but how one feels: The measurement of affect during exercise. Journal of Sport and Exercise Psychology, 11, 304-317.

Haskell, W. L., Yee, M.C., Evans, A., & Irby, P. J. (1993). Simultaneous measurement of heart rate and body motion to quantitate physical activity. Medicine and Science in Sports and Exercise, 25, 109-115.

*Hassmen, P. (1990). Perceptual and physiological responses to cycling and running in groups of trained and untrained subjects. European Journal of Applied Physiology and Occupational Physiology, 60, 445-451.

Hassmen, P., & Koivula, N. (1996). Ratings of perceived exertion by a group of women: Does Type A behavior act as a modifier? Perceptual and Motor Skills, 83, 675-686.

Hedges, L. V. (1981). Distribution theory for Glass' estimator of effect size and related estimators. Journal of Educational Statistics, 6, 107-128.

Hedges, L. V. (1982). Fitting categorical models to effect sizes from a series of experiments. Journal of Educational Statistics, 7, 119-137.

Hedges, L. V., & Olkin, I. (1985). Statistical methods for meta-analysis. San Diego, CA: Academic Press.

Hetzler, R. K., Seip, R. L., Boutcher, S. H., Pierce, E., Snead, D., & Weltman, A. (1991). Effect of exercise modality on ratings of perceived exertion at various lactate concentrations. Medicine and Science in Sports and Exercise, 23, 88-92.

*Higgs, S. L., & Robertson, L. A. (1981). Cyclic variations in perceived exertion and physical work capacity in females. Canadian Journal of Applied Sports Science, 6, 191-196.

Hoffman, M. D., Kassay, K. M., Zeni, A. I., & Clifford, P. S. (1996). Does the amount of exercising muscle alter the aerobic demand of dynamic exercise? European Journal of Applied Physiology, 74, 541-547.

Holloszy, J. O. (1990). The roles of exercise in health maintenance and treatment of disease in middle and old age. In M. Kaneko (Ed.), Fitness for the aged, disabled, and industrial worker (pp. 3-8). Champaign, IL: Human Kinetics.

Horstman, D. H., Morgan, W. P., Cymerman, A., & Stokes, J. (1979). Perception of effort during constant work to self-imposed exhaustion. Perceptual and Motor Skills, 48, 1111-1126.

Hunter, J. E., & Schmidt, F. L. (1990). Methods of meta-analysis. Newbury Park, CA: Sage.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). Meta-analysis. Cumulating research findings across studies. Beverly Hills, CA: Sage.

Ibusuki, T., Kondo, T., Soya, H., & Yagi, H. (1990). Physiological characteristics of wheelchair basketball players. In M. Kaneko (Ed.), Fitness for the aged, disabled, and industrial worker (pp. 164-168). Champaign, IL: Human Kinetics.

Jackson, G. B. (1980). Methods for integrative reviews. Review of Educational Research, 50, 438-460.

*Jackson, A., Dishman, R. K., LaCroix, S., Patton, R., & Weinberg, R. (1981). The heart rate, perceived exertion, and pace of the 1.5 mile run. Medicine and Science in Sports and Exercise, 13, 224-228.

*Kamon, E., Pandolf, K., & Cafarelli, E. (1974). The relationship between perceptual information and physiological responses to exercise in the heat. Journal of Human Ergology, 3, 45-54.

*Kay, C., & Shephard, R. J. (1969). On muscle strength and the threshold of anaerobic work. International Zeitschrift fur Angewandte Physiologie, 27, 311-328.

*Kilbom, A. (1971). Physical training with submaximal intensities in women. Scandinavian Journal of Clinical and Laboratory Investigations, 28, 141-161.

King, C. N., & Senn, M. D. (1996). Exercise testing and prescription. Practical recommendations for the sedentary. Sports Medicine, 21, 326-336.

Kobayashi, M., Hirano, T., & Fukunaga, T. (1990). Evaluation of physical fitness in paraplegic wheelchair basketball players. In M. Kaneko (Ed.), Fitness for the aged, disabled, and industrial worker (pp. 160-163). Champaign, IL: Human Kinetics.

*Kolkhorst, F. W., Mittelstadt, S. W., & Dolgener, F. A. (1996). Perceived exertion and blood lactate concentration during graded treadmill running. European Journal of Applied Physiology, 72, 272-277.

Kulik, J. (1983). Book review. Review of Glass, G. V. et al., Meta-analysis in social research (Sage, 1981). Evaluation News, 4, 101-105.

*Kurokawa, T., & Ueda, T. (1992). Validity of ratings of perceived exertion as an index of exercise intensity in swimming training. Annals of Physiology and Anthropology, 11, 277-288.

*Lamb, K. L. (1995). Children's ratings of effort during cyclic ergometry: An examination of the validity of two effort rating scales. Pediatric Exercise Science, 7, 407-421.

*Lamb, K. L. (1996). Exercise regulation during cyclic ergometry using the Children's Effort Rating Table (CERT) and Ratings of Perceived Exertion (RPE) Scales. Pediatric Exercise Science, 8, 337-350.

Liu, N. Y. S., Plowman, S. A., & Looney, M. A. (1992). The reliability and validity of the 20-meter shuttle test in American students 12 to 15 years old. Research Quarterly for Exercise and Sport, 63, 360-365.

*Ljunggren, G., Ceci, R., & Karlsson, J. (1987). Prolonged exercise at a constant work load on a bicycle ergometer: Ratings of perceived exertion and leg aches and pain as well as measurements of blood lactate accumulation and heart rate. International Journal of Sports Medicine, 8, 109-116.

*Lollgen, H., Graham, T., & Sjogaard, G. (1980). Muscle metabolites, force, and perceived exertion bicycling at varying pedal rates. Medicine and Science in Sports and Exercise, 12, 345-351.

*Lollgen, H., Ulmer, H. V., & Nieding, G. V. (1977). Heart rate and perceptual response to exercise with different pedalling speed in normal subjects and patients. European Journal of Applied Physiology and Occupational Physiology, 37, 297-304.

Martin, M. B., & Anshel, M. H. (1995). Effect of self-monitoring strategies and task complexity on motor performance and affect. Journal of Sport and Exercise Psychology, 17, 153-170.

McArdle, W. D., Katch, F. I., & Katch, V. L. (1996). Exercise physiology. Baltimore, MD: Williams & Wilkins.

McAuley, E., & Rudolph, D. L. (1995). Physical activity, aging, and psychological well-being. Journal of Aging and Physical Activity, 3, 67-96.

Messick, S. (1995). Validity of psychological assessment. American Psychologist, 50, 741-749.

Michael, E., & Eckardt, L. (1972). The selection of hard work by trained and non-trained subjects. Medicine and Science in Sports, 4, 107-110.

Mihivec, P. M. (1981). Sensory cues for perceived exertion: A review. Medicine and Science in Sports and Exercise, 13, 150-163.

*Miller, G. D., Bell, R. D., Collis, M. L., & Hoshizaki, T. B. (1985). The relationship between perceived exertion and heart rate of post 50 year-old volunteers in two different walking activities. Journal of Human Movement Studies, 11, 187-195.

Morgan, W. P. (1973). Psychological factors influencing perceived exertion. Medicine and Science in Sports, 5, 97-103.

*Morgan, W. P., & Pollack, M. L. (1977). Psychologic characterization of the elite distance runner. Annals of the New York Academy of Sciences, 301, 382-403.

Moriya, K., & Fukuchi, Y. (1990). Physical fitness as an important element of successful aging. In M. Kaneko (Ed.), Fitness for the aged, disabled, and industrial worker (pp. 18-20). Champaign, IL: Human Kinetics.

Mosteller, F. M., & Bush, R. R. (1954). Selected quantitative techniques, In G. Lindzey (Ed.), Handbook of social psychology (pp. 289-334). Cambridge, MA: Addison-Wesley.

Murray, T. D., Walker, J. L., Jackson, A. S., Morrow, J. R., Eldridge, J. A., & Rainey, D. L. (1993). Validation of a 20-minute steady-state jog as an estimate of peak oxygen uptake in adolescence. Research Quarterly for Exercise and Sport, 64, 75-82.

Netz, Y. (1987). Relationship between exercise and perceived fitness among middle-aged women and men. Perceptual and Motor Skills, 64, 758.

Noble, B. J. (1982). Preface to the symposium on recent advances in the study and clinical use of perceived exertion. Medicine and Science in Sports and Exercise, 14, 376.

Noble, B. J., Borg, G. A. V., Jacobs, I., Ceci, R., & Kaiser, P. (1983). A category-ratio perceived exertion scale: Relationship to blood and muscle lactates and heart rate. Medicine and Science in Sports and Exercise, 15, 523-528.

*Noble, B. J., Metz, K. F., Pandolf, K. B., & Cafarelli, E. (1973). Perceptual responses to exercise: A multiple regression study. Medicine and Science in Sports, 5, 104-109.

Noble, B. J., & Robertson, R. J. (1996). Perceived exertion. Champaign, IL: Human Kinetics.

*O'Neill, M. E., Cooper, K. A., Mills, C. M., Boyce, E. S., & Hunyor, S. N. (1992). Accuracy of Borg's ratings of perceived exertion in the prediction of heart rates during pregnancy. British Journal of Sports Medicine, 26, 121-124.

Pandolf, K. B. (1983). Advances in the study and application of perceived exertion. Exercise and Sports Science Reviews, 11, 118-158.

*Pandolf, K. B., Cafarelli, E., Noble, B. N., & Metz, K. F. (1972). Perceptual responses during prolonged work. Perceptual and Motor Skills, 35, 975-985.

Pandolf, K. B., & Noble, B. J. (1973). The effect of pedalling speed and resistance changes on perceived exertion for equivalent power outputs on the bicycle ergometer. Medicine and Science in Sports, 5, 132-136.

Parfitt, G., & Eston, R. (1995). Changes in ratings of perceived exertion and psychological affect in the early stages of exercise. Perceptual and Motor Skills, 80, 259-266.

Parfitt, G., Eston, R., & Connolly, D. (1996). Psychological affect at different ratings of perceived exertion in high- and low-active women: A study using a production protocol. Perceptual and Motor Skills, 82, 1035-1042.

Parfitt, G., Markland, D., & Holmes, C. (1994). Responses to physical exertion in active and inactive males and females. Journal of Sport and Exercise Psychology, 16, 178-186.

*Pedersen, P. K., & Welch, H. G. (1977). Oxygen breathing, selected physiological variables and perception of effort during submaximal exercise. In G. Borg (Ed.), Physical work and effort (pp. 385-399). Oxford: Pergamon Press.

Pivarnik, J. M., & Senay, L. C. (1986). Effect of endurance training and heat acclimation on perceived exertion during exercise. Journal of Cardiopulmonary Rehabilitation, 6, 499-504.

Plante, T. G., & Rodin, J. (1990). Physical fitness and enhanced psychological health. Current Psychology Research and Reviews, 9, 3-24.

Potteiger, J. A., & Evans, B. W. (1995). Using heart rate and ratings of perceived exertion to monitor intensity in runners. Journal of Sports Medicine and Physical Fitness, 35, 181-186.

Prusaczyk, W. K., Cureton, K. J., Graham, R. E., & Ray, C. A. (1992). Differential effects of dietary carbohydrate on RPE at the lactate and ventilatory thresholds. Medicine and Science in Sports and Exercise, 24, 568-575.

Rakowski, W., & Mor, V. (1992). The association of physical activity with mortality among older adults in the Longitudinal Study of Aging (1984-1988). Journal of Gerontology: Medical Sciences, 47, M122-M129.

Rejeski, W. J., Hardy, C. J., & Shaw, J. (1991). Psychometric confounds of assessing state anxiety in conjunction with acute bouts of vigorous exercise. Journal of Sport and Exercise Psychology, 13, 65-74.

Robertson, R. J. (1982). Central signals of perceived exertion during dynamic exercise. Medicine and Science in Sports and Exercise, 14, 390-396.

*Robertson, R. J., Falkel, J. E., Drash, A. L., Swank, A. M., Metz, K. F., Spungen, S. A., & LeBoeuf, J. R. (1986). Effect of blood pH on peripheral and central signals of perceived exertion. Medicine and Science in Sports and Exercise, 18, 114-122.

*Robertson, R., Gillespie, R. L., McCarthy, J., & Rose, K. D. (1979). Differentiated perceptions of exertion: Part II. Relationship to local and central physiological responses. Perceptual and Motor Skills, 49, 691-697.

Robertson, R., Goss, F., Michael, T., Moyna, N., Gordon, P., Visich, P., Kang, J., Angelopoulos, T., Dasilva, S., & Metz, K. (1996). Validity of the Borg perceived exertion scale for using semirecumbent ergometry during immersion in water. Perceptual and Motor Skills, 83, 3-13.

Rogosa, D., & Ghandour, G. (1991). Reply to discussants: Statistical models for behavioral observations. Journal of Educational Statistics, 16, 281-294.

Rosenthal, R. (1978). Combining results of independent studies. Psychological Bulletin, 85, 185-193.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. Psychological Bulletin, 86, 638-641.

Rosenthal, R. (1984). Meta-analysis procedures for social research. Beverly Hills, CA: Sage.

Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. The Behavioral and Brain Sciences, 3, 377-386.

*Rudolph, D. L., & McAuley, E. (1996). Self-efficacy and perceptions of effort: A reciprocal relationship. Journal of Sport and Exercise Psychology, 18, 216-223.

Safrit, M. J. (1989). Criterion-referenced measurement: Validity. In M. J. Safrit & T. M. Wood (Eds.), Measurement concepts in physical education and exercise science (pp. 119-135). Champaign, IL: Human Kinetics.

*Sargeant, A. J., & Davies, C. T. M. (1973). Perceived exertion during rhythmic exercise involving different muscle masses. Journal of Human Ergology, 2, 3-11.

*Sargeant, A. J., & Davies, C. T. M. (1977). Perceived exertion of dynamic exercise in normal subjects and patients following leg injury. In G. Borg (Ed.), Physical work and effort (pp. 345-355). Oxford: Pergamon Press.

Seip, R. L., Snead, D., Pierce, E. F., Stein, P., & Weltman, A. (1991). Perceptual responses and blood lactate concentration: Effect of training state. Medicine and Science in Sports and Exercise, 23, 80-87.

Siconolfi, S. F., Lasater, T. M., Snow, R. C. K., & Carleton, R. A. (1985). Self-reported physical activity compared with maximal oxygen uptake. American Journal of Epidemiology, 122, 101-105.

Sidney, K. H., & Shephard, R. J. (1977). Perception of exertion in the elderly, effects of aging, mode of exercise and physical training. Perceptual and Motor Skills, 44, 999-1010.

Sim, J., & Arnell, P. (1993). Measurement validity in physical therapy research. Physical Therapy, 73, 102-115.

*Skinner, J. S., Borg, G., & Buskirk, E. R. (1969). Physiological and perceptual reactions to exertion of young men differing in activity and body size. In D. Franks (Ed.), Proceedings of Symposia on Exercise and Fitness (pp. 53-66). Chicago: Athletic Institute.

Skinner, J. S., Hustler, R., Bergsteinova, V., & Buskirk, E. R. (1973a). The validity and reliability of a rating scale of perceived exertion. Medicine and Science in Sports, 5, 94-96.

Skinner, J. S., Hustler, R., Bergsteinova, V., & Buskirk, E. R. (1973b). Perception of effort during different types of exercise and under different environmental conditions. Medicine and Science in Sports, 5, 110-115.

*Skrinar, G. S., Ingram, S. P., & Pandolf, K. B. (1983). Effect of endurance training on perceived exertion and stress hormones in women. Perceptual and Motor Skills, 57, 1239-1250.

Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. Educational Researcher, 15, 5-11.

*Smith, B. A. (1994). Use of rating of perceived exertion for physical conditioning: A three month validation study in the field. Unpublished masters thesis, Indiana University, Bloomington.

Smolander, J., Korhonen, O., & Ilmarinen, R. (1990). Responses of young and older men during prolonged exercise in dry and humid heat. European Journal of Applied Physiology, 61, 413-418.

Stamford, B. A., & Noble, B. J. (1974). Metabolic cost and perception of effort during bicycle ergometer work performance. Medicine and Science in Sports, 6, 226-231.

*Steed, J., Gaesser, G. A., & Weltman, A. (1994). Rating of perceived exertion and blood lactate concentration during submaximal running. Medicine and Science in Sports and Exercise, 26, 797-803.

*Stevenson, L. A., Kolka, M. A., & Wilkerson, J. E. (1982). Perceived exertion and anaerobic threshold during the menstrual cycle. Medicine and Science in Sports and Exercise, 14, 218-222.

Stoudemire, N. M., Wideman, L., Pass, K. A., McGinnes, C. L., Gaesser, G. A., & Weltman, A. (1996). The validity of regulating blood lactate concentration during running by ratings of perceived exertion. Medicine and Science in Sports and Exercise, 28, 490-495.

Takeshima, N., Kobayashi, F., Watanabe, T., Tanaka, K., Tomita, M., & Pollack, M. L. (1996). Cardiorespiratory responses to cycling exercise in trained and untrained healthy elderly: With special reference to the lactate threshold. Applied Human Science, 15, 267-273.

*Toner, M. M., Drolet, L. L., & Pandolf, K. B. (1986). Perceptual and physiological responses during exercise in cool and cold water. Perceptual and Motor Skills, 62, 211-220.

*Travlos, A. K., & Marisi, D. Q. (1996). Perceived exertion during physical exercise among individuals high and low in fitness. Perceptual and Motor Skills, 82, 419-424.

Trine, M. R., & Morgan, W. P. (1995). Influence of time of day on psychological responses to exercise: A review. Sports Medicine, 20, 328-337.

*Turkulin, K., Zamlic, B., & Pegan, U. (1977). Exercise performance and perceived exertion in patients after myocardial infarction. In G. Borg (Ed.), Physical work and effort (pp. 357-366). Oxford: Pergamon Press.

*Ueda, T., & Kurokawa, T. (1991). Validity of heart rate and ratings of perceived exertion as indices of exercise intensity in a group of children while swimming. European Journal of Applied Physiology and Occupational Physiology, 63, 200-204.

*Ueda, T., Kurokawa, T., Kikkawa, K., & Choi, T. H. (1993). Contribution of differentiated ratings of perceived exertion to overall exertion in women while swimming. European Journal of Applied Physiology and Occupational Physiology, 66, 196-201.

*Ulmer, H.-V., Janz, U., & Lollgen, H. (1977). Aspects of the validity of Borg's Scale. Is it measuring stress or strain? In G. Borg (Ed.), Physical work and effort (pp. 181-196). Oxford: Pergamon Press.

*Williams, J. G., & Eston, R. G. (1986). Does personality influence the perception of effort? The results from a study of secondary school boys. Physical Education Review, 9, 94-99.

*Wilmore, J. H., Roby, F. B., Stanforth, P.R., Buono, M. J., Constable, S. H., Tsao, Y., & Lowdon, B. J. (1985). Ratings of perceived exertion, heart rate, and treadmill speed in the prediction of maximal oxygen uptake during submaximal treadmill exercise. Journal of Cardiopulmonary Rehabilitation, 5, 540-546.

*Winborn, M. D., Meyers, A. W., & Mulling, C. (1988). The effects of gender and experience on perceived exertion. Journal of Sport and Exercise Psychology, 10, 22-31.

Winer, B. J. (1971). Statistical principles in experimental design. New York: McGraw-Hill.

Wolf, F. M. (1986). Meta-analysis. Quantitative methods for research synthesis. Newbury Park, CA: Sage.

Wolf, F. M., & Spies, C. J. (1981). Assessing the consistency of cross-lagged panel effects with the Fisher Combined Test. Proceedings of the American Statistical Association Social Statistics Section, 24, 506-511.

Wong, D. G., Cunningham, D. A., Rechnitzer, P. A., & Howard, J. H. (1990). Effect of an exercise program on the perception of exertion in males at retirement. Canadian Journal of Sports Science, 15, 249-253.

Worthen, B. R., Borg, W. R., & White, K. R. (1993). Measurement and evaluation in the schools. New York: Longman.

*Zeni, A. I., Hoffman, M. D., & Clifford, P. S. (1996). Relationships among heart rate, lactate concentration, and perceived effort for different types of rhythmic exercise in women. Archives of Physical and Medical Rehabilitation, 77, 237-241.

APPENDICES

Table A.1

Coding of Characteristics and Validity Coefficients for Ratings of Perceived Exertion and Heart Rate

Study	n	Gender	Fitness level	RPE scale	Exercise type	Exercise protocol	Study quality	r
Anshel et al. (1991)	39	Female	NS[a]	15-point	Bicycle ergometer	1-level max. exertion	Good	.80
	20	Female	NS	15-point	Bicycle ergometer	1-level max. exertion	Good	.85
	19	Female	NS	15-point	Bicycle ergometer	1-level max. exertion	Good	.75
	20	Female	NS	15-point	Bicycle ergometer	1-level max. exertion	Good	.37
	20	Female	NS	15-point	Bicycle ergometer	1-level max. exertion	Good	.45
	20	Female	NS	15-point	Bicycle ergometer	1-level max. exertion	Good	.64
	19	Female	NS	15-point	Bicycle ergometer	1-level max. exertion	Good	.03
	19	Female	NS	15-point	Bicycle ergometer	1-level max. exertion	Good	.10
	19	Female	NS	15-point	Bicycle ergometer	1-level max. exertion	Good	.33
Arstila et al. (1974)	10	Male	Healthy-inactive	15-point	Bicycle ergometer	Progressive continuous	Excellent	.83
	15	Male	Healthy-inactive	15-point	Bicycle ergometer	Progressive continuous	Excellent	.87
	9	Male	Healthy-inactive	15-point	Bicycle ergometer	Progressive continuous	Excellent	.88
Bacharach (1984)	9	Female	NP[b]	15-point	Bicycle ergometer	Progressive continuous	Cannot judge	.64
	11	Female	NP	15-point	Bicycle ergometer	Progressive continuous	Cannot judge	.65
Bar-Or et al. (1972)	51	Male	Active	15-point	Bicycle ergometer	1-level submax. exertion	Good	.77
	51	Male	Active	15-point	Treadmill	1-level submax. exertion	Good	.80
	19	Male	Sedentary	15-point	Bicycle ergometer	1-level submax. exertion	Good	.77
	19	Male	Sedentary	15-point	Treadmill	1-level submax. exertion	Good	.80

(table continues)

Study	n	Gender	Fitness level	RPE scale	Exercise type	Exercise protocol	Study quality	r
Bloem et al. (1991)	13	Male	Healthy-inactive	Category ratio	Treadmill	Progressive continuous	Good	.35
Borg (1973)	69	Male	Healthy-inactive	15-point	Bicycle ergometer	1-level submax. exertion	Excellent	.62
	63	Male	Healthy-inactive	21-point	Bicycle ergometer	1-level submax. exertion	Excellent	.56
	43	Male	Healthy-inactive	15-point	Bicycle ergometer	Progressive continuous	Excellent	.72
	43	Male	Healthy-inactive	21-point	Bicycle ergometer	Progressive continuous	Excellent	.60
	63	Male	Healthy-inactive	9-point	Bicycle ergometer	1-level submax. exertion	Excellent	.54
	46	Male	Healthy-inactive	9-point	Bicycle ergometer	Progressive continuous	Excellent	.52
Borg et al. (1977)	20	Male	Active	15-point	Bicycle ergometer	1-level submax. exertion	Excellent	.82
Borg et al. (1985)	28	Male	Active	Category-ratio	Bicycle ergometer	Progressive continuous	Excellent	.28
	28	Male	Active	Category-ratio	Bicycle ergometer	Progressive continuous	Excellent	.18
	28	Male	Active	Category-ratio	Bicycle ergometer	Progressive continuous	Excellent	.11
	28	Male	Active	Category-ratio	Bicycle ergometer	Progressive continuous	Excellent	.41
	28	Male	Active	Category-ratio	Bicycle ergometer	Progressive continuous	Excellent	.61
	28	Male	Active	Category-ratio	Bicycle ergometer	Progressive continuous	Excellent	.58
	28	Male	Active	Category-ratio	Bicycle ergometer	Progressive continuous	Excellent	.56
	28	Male	Active	Category-ratio	Bicycle ergometer	Progressive continuous	Excellent	.91
Borg et al. (1972)	70	Male	NP	15-point	Bicycle ergometer	1-level submax. exertion	Poor	.24
	70	Male	NP	15-point	Bicycle ergometer	1-level submax. exertion	Poor	.26
	70	Male	NP	15-point	Treadmill	1-level submax. exertion	Poor	.25
	70	Male	NP	15-point	Treadmill	1-level submax. exertion	Poor	.34

(table continues)

Study	n	Gender	Fitness Level	RPE Scale	Exercise Type	Exercise Protocol	Study Quality	r
Borg, Van den	10	Male	Active	15-point	Treadmill	1-level submax. exertion	Excellent	.99
Burg et al.	10	Male	Active	15-point	Bicycle ergometer	1-level submax. exertion	Excellent	.99
(1987)	10	Male	Active	15-point	Track running	1-level submax. exertion	Excellent	.99
Brown et al. (1996b)	24	NS	Active	15-point	Swimming	1-level submax. exertion	Excellent	.81
Butts (1982)	127	Female	NP	15-point	Treadmill	Progressive continuous	Good	.46
	127	Female	NP	15-point	Treadmill	Progressive continuous	Good	.46
	127	Female	NP	15-point	Treadmill	Progressive continuous	Good	.63
	127	Female	NP	15-point	Treadmill	Progressive continuous	Good	.74
	106	Female	NP	15-point	Treadmill	Progressive continuous	Good	.62
	43	Female	NP	15-point	Treadmill	Progressive continuous	Good	.25
Clapp &	36	Female	Highly fit	15-point	Track running	1-level submax. exertion	Poor	.74
Little (1994)	53	Female	Active	15-point	Track running	1-level submax. exertion	Poor	.14
Costa & Gaffuri (1977)	12	NS	Healthy-inactive	15-point	Bicycle ergometer	Progressive continuous	Poor	.84
Duncan et al.	10	Male	Healthy-inactive	15-point	Treadmill	Progressive continuous	Excellent	.98
(1996)	10	Male	Healthy-inactive	15-point	Bicycle ergometer	Progressive continuous	Excellent	.98
Edwards et al.	3	Male	NP	15-point	Bicycle ergometer	Progressive continuous	Excellent	.88
(1972)	3	Male	NP	15-point	Bicycle ergometer	Progressive intermittent	Excellent	.86

(table continues)

Study	n	Gender	Fitness level	RPE scale	Exercise type	Exercise protocol	Study quality	r
Gamberale (1972)	12	Male	Healthy-inactive	15-point	Bicycle ergometer	Progressive continuous	Excellent	.53
	12	Male	Healthy-inactive	15-point	Bicycle ergometer	Progressive continuous	Excellent	.57
	12	Male	Healthy-inactive	15-point	Bicycle ergometer	Progressive continuous	Excellent	.72
	12	Male	Healthy-inactive	15-point	Bicycle ergometer	Progressive continuous	Excellent	.47
Goslin & Rorke (1986)	10	Male	Highly fit	Category-ratio	Treadmill	Random intermittent	Good	.47
Hardy & Rejeski (1989)	30	Male	Active	15-point	Bicycle ergometer	Progressive continuous	Good	.85
Hassmen (1990)	6	Male	Sedentary	15-point	Treadmill	1-level max. exertion	Excellent	.99
	6	Male	Sedentary	15-point	Bicycle ergometer	1-level max. exertion	Excellent	.99
	6	Male	Highly fit	15-point	Treadmill	1-level max. exertion	Excellent	.99
	6	Male	Highly fit	15-point	Bicycle ergometer	1-level max. exertion	Excellent	.99
	6	Male	Highly fit	15-point	Treadmill	1-level max. exertion	Excellent	.99
	6	Male	Highly fit	15-point	Bicycle ergometer	1-level max. exertion	Excellent	.99
	6	Male	Highly fit	15-point	Treadmill	1-level max. exertion	Excellent	.99
	6	Male	Highly fit	15-point	Bicycle ergometer	1-level max. exertion	Excellent	.99
Higgs & Robertson (1981)	12	Female	Active	21-point	Treadmill	1-level submax. exertion	Excellent	.75
	12	Female	Active	21-point	Treadmill	1-level submax. exertion	Excellent	.61
	12	Female	Active	21-point	Treadmill	1-level submax. exertion	Excellent	.54
	12	Female	Active	21-point	Treadmill	1-level submax. exertion	Excellent	.78

(table continues)

Study	n	Gender	Fitness level	RPE scale	Exercise type	Exercise protocol	Study quality	r
Jackson et al. (1981)	67	Male	Active	15-point	Track-running	1-level submax. exertion	Good	.16
Kamon et al. (1974)	10	Male	NP	15-point	Bicycle ergometer	Progressive continuous	Excellent	.48
Kilbom (1971)	12	Female	Active	15-point	Bicycle ergometer	Progressive intermittent	Excellent	.90
	8	Female	Active	15-point	Bicycle ergometer	Progressive intermittent	Excellent	.88
	13	Female	Active	15-point	Bicycle ergometer	Progressive intermittent	Excellent	.82
	10	Female	Active	15-point	Bicycle ergometer	Progressive intermittent	Excellent	.87
	8	Female	Active	15-point	Bicycle ergometer	Progressive intermittent	Excellent	.88
	12	Female	Active	15-point	Bicycle ergometer	Progressive intermittent	Excellent	.80
Kolkhorst et al. (1996)	10	NP	Healthy-inactive	15-point	Treadmill	Progressive intermittent	Excellent	.73
	10	NP	Healthy-inactive	15-point	Treadmill	Progressive intermittent	Excellent	.48
Kurakawa & Ueda (1992)	10	Female	Active	15-point	Swimming	Progressive intermittent	Excellent	.96
	7	Male	Active	15-point	Swimming	Progressive intermittent	Excellent	.97
	8	NP	Active	15-point	Swimming	1-level submax. exertion	Excellent	.46
	8	NP	Active	15-point	Swimming	1-level submax. exertion	Excellent	.85
	8	NP	Active	15-point	Swimming	1-level submax. exertion	Excellent	.85
	8	NP	Active	15-point	Swimming	1-level submax. exertion	Excellent	.24
	8	NP	Active	15-point	Swimming	1-level submax. exertion	Excellent	.77
	8	NP	Active	15-point	Swimming	1-level submax. exertion	Excellent	.82

(table continues)

Study	n	Gender	Fitness level	RPE scale	Exercise type	Exercise protocol	Study quality	r
Lamb (1995)	34	NP	Healthy-inactive	15-point	Bicycle ergometer	1-level submax. exertion	Good	.57
Lamb (1996)	33	NP	Healthy-inactive	15-point	Bicycle ergometer	Progressive intermittent	Excellent	.58
	12	Male	Healthy-inactive	15-point	Bicycle ergometer	Progressive intermittent	Excellent	.73
	21	Female	Healthy-inactive	15-point	Bicycle ergometer	Progressive intermittent	Excellent	.52
Lollgen et al. (1977)	128	Male	Healthy-inactive	15-point	Bicycle ergometer	Progressive intermittent	Good	.63
Miller et al. (1985)	35	Male	Nealthy-inactive	15-point	Treadmill	1-level submax. exertion	Excellent	.17
	35	Male	Healthy-inactive	15-point	Treadmill	1-level submax. exertion	Excellent	.25
	35	Female	Healthy-inactive	15-point	Treadmill	1-level submax. exertion	Excellent	.48
	35	Female	Healthy-inactive	15-point	Treadmill	1-level submax. exertion	Excellent	.43
	35	Male	Healthy-inactive	15-point	Treadmill	1-level submax. exertion	Excellent	.23
	35	Female	Healthy-inactive	15-point	Treadmill	1-level submax. exertion	Excellent	.47
Morgan & Pollack (1977)	27	Male	Highly fit	15-point	Treadmill	1-level submax. exertion	Excellent	.43
Noble et al. (1973)	6	Male	NP	15-point	Bicycle ergometer	1-level submax. exertion	Cannot judge	.81
	6	Male	NP	15-point	Bicycle ergometer	1-level submax. exertion	Cannot judge	.85
	6	Male	NP	15-point	Bicycle ergometer	1-level submax. exertion	Cannot judge	.70
	6	Male	NP	15-point	Bicycle ergometer	1-level submax. exertion	Cannot judge	.75
	6	Male	NP	15-point	Bicycle ergometer	1-level submax. exertion	Cannot judge	.92
	6	Male	NP	15-point	Bicycle ergometer	1-level submax. exertion	Cannot judge	.63

(table continues)

Study	n	Gender	Fitness level	RPE scale	Exercise type	Exercise protocol	Study quality	r
O'Neill et al. (1992)	11	Female	Healthy-inactive	15-point	Treadmill	Progressive continuous	Good	.83
	12	Female	Healthy-inactive	15-point	Bicycle ergometer	1-level submax. exertion	Good	.74
	24	Female	Healthy-inactive	15-point	NP	1-level submax. exertion	Good	.39
Pedersen & Welch (1977)	6	Male	Healthy-inactive	15-point	Bicycle ergometer	Progressive continuous	Excellent	.81
Sargeant & Davies (1977)	12	NS	Healthy-inactive	15-point	Bicycle ergometer	Progressive continuous	Excellent	.87
Sargeant & Davies (1973)	6	Male	NP	15-point	Bicycle ergometer	Progressive continuous	Excellent	.84
Skinner et al. (1969)	26	Male	NP	15-point	Bicycle ergometer	Progressive continuous	Cannot judge	.83
Smith (1994)	18	NS	Sedentary	15-point	Treadmill	Progressive continuous	Poor	.03
Stevenson et al. (1982)	6	Female	NP	9-point	Bicycle ergometer	1-level submax. exertion	Good	.87
Toner et al. (1986)	8	Male	Healthy-inactive	15-point	Bicycle ergometer	Progressive intermittent	Good	.68

(table continues)

Study	n	Gender	Fitness level	RPE scale	Exercise type	Exercise protocol	Study quality	r
Travlos &	10	Male	Highly fit	15-point	Bicycle ergometer	Progressive continuous	Excellent	.41
Marisi (1996)	10	Male	Highly fit	15-point	Bicycle ergometer	Progressive continuous	Excellent	.63
	10	Male	Highly fit	15-point	Bicycle ergometer	Progressive continuous	Excellent	.59
	10	Male	Highly fit	15-point	Bicycle ergometer	Progressive continuous	Excellent	.63
	10	Male	Highly fit	15-point	Bicycle ergometer	Progressive continuous	Excellent	.50
	10	Male	Sedentary	15-point	Bicycle ergometer	Progressive continuous	Excellent	.25
	10	Male	Sedentary	15-point	Bicycle ergometer	Progressive continuous	Excellent	.23
	10	Male	Sedentary	15-point	Bicycle ergometer	Progressive continuous	Excellent	.59
	10	Male	Sedentary	15-point	Bicycle ergometer	Progressive continuous	Excellent	.67
	10	Male	Sedentary	15-point	Bicycle ergometer	Progressive continuous	Excellent	.57
	10	Male	NS	15-point	Bicycle ergometer	Progressive continuous	Excellent	.37
	10	Male	NS	15-point	Bicycle ergometer	Progressive continuous	Excellent	.52
	10	Male	NS	15-point	Bicycle ergometer	Progressive continuous	Excellent	.61
	10	Male	NS	15-point	Bicycle ergometer	Progressive continuous	Excellent	.64
	10	Male	NS	15-point	Bicycle ergometer	Progressive continuous	Excellent	.54
Turkulin et al.	17	Male	Sedentary	15-point	Bicycle ergometer	1-Level max. exertion	Excellent	.59
(1977)	22	Male	Active	15-point	Bicycle ergometer	1-Level max. exertion	Excellent	.69
Ueda et al. (1993)	10	Female	Active	15-point	Swimming	Progressive intermittent	Excellent	.87
Ueda & Kurakawa (1991)	6	Male	Healthy-inactive	15-point	Swimming	Progressive intermittent	Excellent	.88

(table continues)

Study	n	Gender	Fitness level	RPE scale	Exercise type	Exercise protocol	Study quality	r
Ulmer et al. (1977)	60	NS	Highly fit	15-point	Bicycle ergometer	Random intermittent	Good	.89
Williams & Eston (1988)	30	Male	Healthy-inactive	15-point	Bicycle ergometer	Progressive continuous	Good	.74
Wilmore et al. (1985)	42	NS	Healthy-inactive	15-point	Treadmill	1-level submax. exertion	Excellent	.32
	42	NS	Healthy-inactive	15-point	Treadmill	1-level submax. exertion	Excellent	.29
	42	NS	Healthy-inactive	15-point	Treadmill	1-level submax. exertion	Excellent	.27
Winborn et al. (1988)	12	Male	Highly fit	15-point	Bicycle ergometer	Random intermittent	Excellent	.58
	12	Male	Highly fit	15-point	Bicycle ergometer	Random intermittent	Excellent	.64
	12	Male	Highly fit	15-point	Bicycle ergometer	Random intermittent	Excellent	.61
	12	Male	Sedentary	15-point	Bicycle ergometer	Random intermittent	Excellent	-.31
	12	Male	Sedentary	15-point	Bicycle ergometer	Random intermittent	Excellent	.30
	12	Male	Sedentary	15-point	Bicycle ergometer	Random intermittent	Excellent	.50
	12	Female	Highly fit	15-point	Bicycle ergometer	Random intermittent	Excellent	.47
	12	Female	Highly fit	15-point	Bicycle ergometer	Random intermittent	Excellent	.01
	12	Female	Highly fit	15-point	Bicycle ergometer	Random intermittent	Excellent	.66
	12	Female	Sedentary	15-point	Bicycle ergometer	Random intermittent	Excellent	-.30
	12	Female	Sedentary	15-point	Bicycle ergometer	Random intermittent	Excellent	.53
	12	Female	Sedentary	15-point	Bicycle ergometer	Random intermittent	Excellent	.01
	12	Male	Highly fit	15-point	Bicycle ergometer	Random intermittent	Excellent	.61
	12	Male	Sedentary	15-point	Bicycle ergometer	Random intermittent	Excellent	.18

(table continues)

Study	n	Gender	Fitness level	RPE scale	Exercise type	Exercise protocol	Study quality	r
Winborn et al.	12	Female	Highly fit	15-point	Bicycle ergometer	Random intermittent	Excellent	.41
(1988)	12	Female	Sedentary	15-point	Bicycle ergometer	Random intermittent	Excellent	.10
Zeni et al.	10	Female	Healthy-inactive	15-point	Treadmill	Progressive continuous	Excellent	.99
(1996)	10	Female	Healthy-inactive	15-point	Bicycle ergometer	Progressive continuous	Excellent	.99

[a] NP is not provided.
[b] NS is not specified.

Table A.2

Coding of Characteristics and Validity Coefficients for Ratings of Perceived Exertion and Blood Lactate

Study	n	Gender	Fitness level	RPE scale	Exercise type	Exercise protocol	Study quality	r
Allen & Pandolf (1977)	12	Male	NP[a]	15-point	Treadmill	1-level submax. exertion	Good	.64
Bloem et al. (1991)	13	Male	Healthy-inactive	Category-ratio	Treadmill	Progressive continuous	Good	.61
Borg et al. (1985)	28	Male	Active	Category-ratio	Bicycle ergometer	Progressive continuous	Excellent	.45
	28	Male	Active	Category-ratio	Bicycle ergometer	Progressive continuous	Excellent	.18
	28	Male	Active	Category-ratio	Bicycle ergometer	Progressive continuous	Excellent	.34
	28	Male	Active	Category-ratio	Bicycle ergometer	Progressive continuous	Excellent	.31
	28	Male	Active	Category-ratio	Bicycle ergometer	Progressive continuous	Excellent	.53
	28	Male	Active	Category-ratio	Bicycle ergometer	Progressive continuous	Excellent	.40
	28	Male	Active	Category-ratio	Bicycle ergometer	Progressive continuous	Excellent	.08
	28	Male	Active	Category-ratio	Bicycle ergometer	Progressive continuous	Excellent	.81
Demello et al. (1986)	10	Male	Active	15-point	Treadmill	Progressive continuous	Excellent	.67
	10	Female	Active	15-point	Treadmill	Progressive continuous	Excellent	.69
	10	Male	Healthy-inactive	15-point	Treadmill	Progressive continuous	Excellent	.74
	10	Female	Healthy-inactive	15-point	Treadmill	Progressive continuous	Excellent	.84

(table continues)

Study	n	Gender	Fitness level	RPE scale	Exercise type	Exercise protocol	Study quality	r
Edwards et al. (1987)	3	Male	NS[b]	15-point	Bicycle ergometer	Progressive continuous	Excellent	.77
	3	Male	NS	15-point	Bicycle ergometer	Progressive intermittent	Excellent	.63
Kay & Shepard (1969)	10	Male	NP	15-point	Bicycle ergometer	1-level submax. exertion	Good	.15
Kilbom (1971)	12	Female	Active	15-point	Bicycle ergometer	Progressive intermittent	Excellent	.89
	8	Female	Active	15-point	Bicycle ergometer	Progressive intermittent	Excellent	.72
	13	Female	Active	15-point	Bicycle ergometer	Progressive intermittent	Excellent	.75
	8	Female	Active	15-point	Bicycle ergometer	Progressive intermittent	Excellent	.62
	12	Female	Active	15-point	Bicycle ergometer	Progressive intermittent	Excellent	.71
	12	Female	Active	15-point	Bicycle ergometer	Progressive intermittent	Excellent	.74
Ljunggren et al. (1988)	22	Male	Healthy-inactive	Category-ratio	Bicycle ergometer	1-level max. exertion	Excellent	.42
	22	Male	Healthy-inactive	Category-ratio	Bicycle ergometer	1-level max. exertion	Excellent	.54
Morgan & Pollack (1977)	27	Male	Highly fit	15-point	Treadmill	1-level submax. exertion	Excellent	.61
Pedersen & Welch (1977)	6	Male	Healthy-inactive	15-point	Bicycle ergometer	Progressive continuous	Excellent	.80

(table continues)

Study	n	Gender	Fitness level	RPE scale	Exercise type	Exercise protocol	Study quality	r
Robertson et al. (1989)	50	Male	Active	9-point	Bicycle ergometer	Random intermittent	Excellent	.45
	50	Male	Active	9-point	Bicycle ergometer	Random intermittent	Excellent	.45
	50	Male	Active	9-point	Bicycle ergometer	Random intermittent	Excellent	.47
Skrinar et al. (1983)	15	Female	Active	15-point	Treadmill	Progressive intermittent	Good	.68
Steed et al. (1994)	9	Male	Active	15-point	Treadmill	1-level submax. exertion	Good	.90
Ueda et al. (1990)	10	Female	Active	15-point	Swimming	Progressive intermittent	Excellent	.73
	10	Female	Active	15-point	Swimming	Progressive intermittent	Excellent	.78
Zeni et al. (1996)	10	Female	Healthy-inactive	15-point	Treadmill	Progressive continuous	Excellent	.99
	10	Female	Healthy-inactive	15-point	Bicycle ergometer	Progressive continuous	Excellent	.99

[a] NP is not provided.
[b] NS is not specified.

Table A.3

Coding of Characteristics and Validity Coefficients for Ratings of Perceived Exertion and Oxygen Uptake

Study	n	Gender	Fitness level	RPE scale	Exercise type	Exercise protocol	Study quality	r
Bloem et al. (1991)	52	Male	Healthy-inactive	Category-ratio	Treadmill	Progressive continuous	Good	.39[d]
	52	Male	Healthy-inactive	Category-ratio	Treadmill	Progressive continuous	Good	.60[d]
	20	Male	Healthy-inactive	Category-ratio	Treadmill	Progressive continuous	Good	.57[d]
	24	Male	Healthy-inactive	Category-ratio	Treadmill	Progressive continuous	Good	.68[d]
Brown et al. (1996b)	12	Male	Active	15-point	Swimming	Progressive continuous	Excellent	.82[d]
	12	Female	Active	15-point	Swimming	Progressive continuous	Excellent	.77[d]
	24	NS[a]	Active	15-point	Swimming	Progressive continuous	Excellent	.89[d]
Butts (1982)	127	Female	Active	15-point	Bicycle ergom.[b]	Progressive continuous	Good	-.21[d]
	127	Female	Active	15-point	Bicycle ergom.	Progressive continuous	Good	.12[f]
Demello et al. (1987)	10	Male	Active	15-point	Treadmill	Progressive continuous	Excellent	.62[f]
	10	Female	Active	15-point	Treadmill	Progressive continuous	Excellent	.75[f]
	10	Male	Healthy-inactive	15-point	Treadmill	Progressive continuous	Excellent	.61[f]
	10	Female	Healthy-inactive	15-point	Treadmill	Progressive continuous	Excellent	.77[f]

(table continues)

Study	n	Gender	Fitness level	RPE scale	Exercise type	Exercise protocol	Study quality	r
Edwards et al. (1972)	3	Male	NP[c]	15-point	Bicycle ergom.	Progressive continuous	Excellent	.97[e]
	3	Male	NP	15-point	Bicycle ergom.	Progressive intermittent	Excellent	.92[e]
	3	Male	NP	15-point	Bicycle ergom.	Progressive continuous	Excellent	.94[f]
	3	Male	NP	15-point	Bicycle ergom.	Progressive intermittent	Excellent	.90[f]
	3	Male	NP	15-point	Bicycle ergom.	Progressive continuous	Excellent	.67[g]
	3	Male	NP	15-point	Bicycle ergom.	Progressive continuous	Excellent	-.39[g]
Goslin & Rorke (1986)	10	Male	Highly fit	Category-ratio	Treadmill	Random intermittent	Good	.75[e]
Hardy & Rejeski (1989)	30	Male	Active	15-point	Bicycle ergom.	Progressive continuous	Good	.82[e]
	30	Male	Healthy-inactive	15-point	Bicycle ergom.	Progressive continuous	Good	.77[f]
	30	Male	Active	15-point	Bicycle ergom.	Progressive continuous	Good	.73[g]
Kamon et al. (1974)	10	Male	NP	15-point	Bicycle ergom.	Progressive continuous	Excellent	.52[e]
	10	Male	NP	15-point	Bicycle ergom.	Progressive continuous	Excellent	.77[f]
	10	Male	NP	15-point	Bicycle ergom.	Progressive continuous	Excellent	.47[g]
Kolkhorst et al. (1996)	10	NS[a]	Healthy-inactive	15-point	Treadmill	Progressive intermittent	Excellent	.15[e]
	10	NS	Healthy-inactive	15-point	Treadmill	Progressive intermittent	Excellent	.08[e]
	10	NS	Healthy-inactive	15-point	Treadmill	Progressive intermittent	Excellent	.22[f]
	10	NS	Healthy-inactive	15-point	Treadmill	Progressive intermittent	Excellent	.14[f]
Kurakawa & Ueda (1992)	10	Female	Active	15-point	Swimming	Progressive intermittent	Excellent	.91[d]
	7	Male	Active	15-point	Swimming	Progressive intermittent	Excellent	.93[d]

(table continues)

Study	n	Gender	Fitness level	RPE scale	Exercise type	Exercise protocol	Study quality	r
Lollgen et al. (1980)	6	Male	Active	15-point	Bicycle ergom.	Progressive intermittent	Excellent	.79[d]
Morgan & Pollack (1977)	27	Male	Highly fit	15-point	Treadmill	1-level submax. exertion	Excellent	.52[f]
Noble et al. (1973)	6	Male	NS	15-point	Bicycle ergom.	1-level submax. exertion	Excellent	.68[e]
	6	Male	NS	15-point	Bicycle ergom.	1-level submax. exertion	Excellent	.89[e]
	6	Male	NS	15-point	Bicycle ergom.	1-level submax. exertion	Excellent	.68[e]
	6	Male	NS	15-point	Bicycle ergom.	1-level submax. exertion	Excellent	.83[e]
	6	Male	NS	15-point	Bicycle ergom.	1-level submax. exertion	Excellent	.88[e]
	6	Male	NS	15-point	Bicycle ergom.	1-level submax. exertion	Excellent	.91[e]
	6	Male	NS	15-point	Bicycle ergom.	1-level submax. exertion	Excellent	.56[f]
	6	Male	NS	15-point	Bicycle ergom.	1-level submax. exertion	Excellent	.75[f]
	6	Male	NS	15-point	Bicycle ergom.	1-level submax. exertion	Excellent	.56[f]
	6	Male	NS	15-point	Bicycle ergom.	1-level submax. exertion	Excellent	.68[f]
	6	Male	NS	15-point	Bicycle ergom.	1-level submax. exertion	Excellent	.65[f]
	6	Male	NS	15-point	Bicycle ergom.	1-level submax. exertion	Excellent	.76[f]
	6	Male	NS	15-point	Bicycle ergom.	1-level submax. exertion	Excellent	.77[g]
	6	Male	NS	15-point	Bicycle ergom.	1-level submax. exertion	Excellent	.90[g]
	6	Male	NS	15-point	Bicycle ergom.	1-level submax. exertion	Excellent	.65[g]
	6	Male	NS	15-point	Bicycle ergom.	1-level submax. exertion	Excellent	.90[g]
	6	Male	NS	15-point	Bicycle ergom.	1-level submax. exertion	Excellent	.75[g]
	6	Male	NS	15-point	Bicycle ergom.	1-level submax. exertion	Excellent	.48[g]
Pandolf et al. (1972)	10	Male	Healthy-inactive	15-point	Bicycle ergom.	1-level submax. exertion	Good	.78[f]
	10	Male	Healthy-inactive	15-point	Bicycle ergom.	1-level submax. exertion	Good	.60[g]
Pedersen & Welch (1977)	6	Male	Healthy-inactive	15-point	Bicycle ergom.	1-level submax. exertion	Excellent	.78[e]
	6	Male	Healthy-inactive	15-point	Bicycle ergom.	1-level submax. exertion	Excellent	.83[f]

(table continues)

Study	n	Gender	Fitness level	RPE scale	Exercise type	Exercise protocol	Study quality	r
Robertson et al. (1986)	10	Male	Active	15-point	Bicycle ergom.	1-level max. exertion	Good	.89[f]
	10	Male	Active	15-point	Bicycle ergom.	1-level max. exertion	Good	.86[f]
	10	Male	Active	15-point	Bicycle ergom.	1-level max. exertion	Good	.87[f]
Rudolph & McAuley (1996)	50	Male	Active	15-point	Treadmill	1-level max. exertion	Good	-.61[d]
Sargeant & Davies (1977)	12	NS	Healthy-inactive	15-point	Bicycle ergom.	Progressive continuous	Excellent	.86[d]
	12	NS	Healthy-inactive	15-point	Bicycle ergom.	Progressive continuous	Excellent	.77[e]
	12	NS	Healthy-inactive	15-point	Bicycle ergom.	Progressive continuous	Excellent	.81[f]
Sargeant & Davies (1973)	6	Male	NP	15-point	Bicycle ergom.	Progressive continuous	Excellent	.88[d]
	6	Male	NP	15-point	Bicycle ergom.	Progressive continuous	Excellent	.87[e]
	6	Male	NP	15-point	Bicycle ergom.	Progressive continuous	Excellent	.90[f]
Toner et al. (1986)	8	Male	Healthy-inactive	15-point	Bicycle ergom.	Progressive intermittent	Good	.51[e]
	8	Male	Healthy-inactive	15-point	Bicycle ergom.	Progressive intermittent	Good	.61[f]
Ueda et al. (1993)	10	Female	Active	15-point	Swimming	Progressive intermittent	Excellent	.99[d]
	10	Female	Active	15-point	Swimming	Progressive intermittent	Excellent	.67[f]
Ueda & Kurakawa (1991)	6	Male	Healthy-inactive	15-point	Swimming	Progressive intermittent	Excellent	.91[d]
Wilmore et al. (1985)	42	NS	Healthy-inactive	15-point	Treadmill	1-level submax. exertion	Excellent	.36[e]
	42	NS	Healthy-inactive	15-point	Treadmill	1-level submax. exertion	Excellent	.33[e]
	42	NS	Healthy-inactive	15-point	Treadmill	1-level submax. exertion	Excellent	.31[e]

[a] NS is not specified; [b] Bicycle ergometer; [c] NP is not provided; [d] % maximal oxygen uptake; [e] oxygen uptake; [f] minute ventilation; [g] respiratory rate.

Table A.4

Descriptive Statistics for Weighted Correlations for Heart Rate

Study Feature	k[a]	M_1	SD
Overall	143	.532	.253
Gender overall	143	.531	.253
Males	94	.525	.271
Females	49	.538	.226
Fitness level overall	123	.545	.253
Sedentary	19	.379	.365
Healthy-inactive	43	.521	.218
Active	38	.593	.309
Highly fit	23	.660	.230
RPE scale overall	161	.532	.253
15-point	142	.535	.262
21-point	6	.605	.080
9-point	3	.549	.121
Category-ratio	10	.450	.236
Exercise type overall	160	.533	.253
Bicycle ergometer	108	.577	.247
Treadmill	37	.482	.225
Track running	4	.332	.345
Swimming	11	.778	.194
Exercise protocol overall	161	.532	.253
Progressive continuous	58	.582	.200
Progressive intermittent	18	.648	.156
Random intermittent	18	.450	.375
Maximal exertion	10	.841	.198
Submaximal exertion	57	.457	.259
Study quality overall	152	.528	.253
Excellent	112	.530	.267
Good	32	.587	.203
Poor	8	.304	.194

[a] k is the number of correlations.

Table A.5

Descriptive Statistics for Weighted Correlations for Blood Lactate

Study Feature	k[a]	M_1	SD
Overall	36	.530	.229
Gender overall	36	.530	.229
Males	23	.460	.192
Females	13	.778	.108
Fitness level overall	32	.532	.229
Healthy-inactive	8	.674	.215
Active	23	.499	.221
Highly fit	1	.610	.000[b]
RPE scale overall	36	.530	.229
15-point	22	.721	.164
9-point	3	.457	.009
Category-ratio	11	.412	.196
Exercise type overall	36	.530	.229
Bicycle ergometer	24	.471	.218
Treadmill	10	.712	.121
Swimming	2	.755	.025
Exercise protocol overall	36	.530	.229
Progressive continuous	17	.494	.281
Progressive intermittent	10	.735	.071
Random intermittent	3	.457	.009
Maximal exertion	2	.480	.060
Submaximal exertion	4	.582	.221
Study quality overall	36	.530	.229
Excellent	31	.523	.231
Good	5	.600	.224

[a] k is the number of correlations.

[b] Based on one validity coefficient from a simple size of 27 from one study.

Table A.6

Descriptive Statistics for Weighted Correlations for Oxygen Uptake

Study Feature	k[a]	M_1	SD
Overall	74	.427	.465
Gender overall	63	.425	.465
Males	55	.577	.387
Females	8	.123	.603
Fitness level overall	44	.376	.465
Healthy-inactive	24	.509	.217
Active	18	.241	.651
Highly fit	2	.582	.115
RPE scale overall	74	.427	.465
15-point	69	.408	.498
Category-ratio	5	.549	.130
Exercise type overall	74	.427	.465
Bicycle ergometer	48	.432	.501
Treadmill	18	.331	.381
Swimming	5	.858	.091
Exercise protocol overall	74	.427	.465
Progressive continuous	31	.413	.460
Progressive intermittent	14	.565	.346
Random intermittent	1	.750	.000[b]
Maximal exertion	4	-.055	.907
Submaximal exertion	24	.532	.250
Study quality overall	74	.427	.465
Excellent	56	.609	.269
Good	18	.264	.526

[a] k is the number of correlations.

[b] Based on one validity coefficient from a sample size of 10 from one study.

Table A.7

Descriptive Statistics for Statistically Significant Two–Way Interactions by Study Characteristics for Heart Rate

Gender × RPE Scale

		Gender	
		Males	Females
RPE	1	.655 (80)	.554 (44)
	2	.580 (2)	.670 (4)
	3	.530 (2)	.870 (1)
	4	.446 (10)	

Gender × Exercise Type

		Gender	
		Males	Females
Type	1	.622 (74)	.544 (29)
	2	.629 (16)	.603 (15)
	3	.580 (2)	.444 (2)
	5	.920 (2)	.930 (2)

Gender × Protocol

		Gender	
		Males	Females
Protocol	1	.601 (43)	.661 (11)
	2	.793 (6)	.834 (9)
	3	.398 (9)	.236 (8)
	4	.927 (10)	—
	5	.602 (26)	.527 (21)

RPE Scale × Exercise Type

		RPE Scale			
		1	2	3	4
Type	1	.619 (95)	.580 (2)	.643 (3)	.455 (8)
	2	.572 (31)	.670 (4)	—	.410 (2)
	3	.512 (4)	—	—	—
	5	.771 (11)	—	—	—

RPE Scale × Study Quality

		RPE Scale			
		1	2	3	4
Quality	1	.633 (46)	.640 (6)	.530 (2)	.455 (8)
	2	.590 (29)	—	.870 (1)	.410 (2)
	3	.356 (8)	—	—	—

Exercise Type × Protocol

		Exercise Type			
		1	2	3	5
Protocol	1	.620 (46)	.577 (11)	—	.810 (1)
	2	.762 (12)	.605 (2)	—	.921 (4)
	3	.346 (17)	.470 (1)	—	—
	4	.880 (6)	.999 (4)	—	—
	5	.620 (27)	.484 (19)	.512 (4)	.664 (6)

Exercise Type × Study Quality

		Study Quality		
		1	2	3
Type	1	.592 (77)	.623 (19)	.448 (3)
	2	.617 (23)	.583 (11)	.206 (3)
	3	.999 (1)	.160 (1)	.444 (2)
	5	.771 (11)	—	—

Protocol × Study Quality

		Study Quality		
		1	2	3
Protocol	1	.622 (43)	.593 (10)	.436 (2)
	2	.796 (16)	.657 (2)	—
	3	.313 (16)	.679 (2)	—
	4	.927 (10)	—	—
	5	.575 (27)	.566 (18)	.330 (6)

Note. RPE 1 is 15-point scale. RPE 2 is 21-point scale. RPE 3 is 9-point scale.
RPE 4 is Category-ratio scale.
Type 1 is bicycle ergometer. Type 2 is treadmill. Type 3 is track running. Type 5 is swimming.
Protocol 1 is progressive continuous. Protocol 2 is progressive intermittent. Protocol 3 is random intermittent. Protocol 4 is one-level maximal exertion. Protocol 5 is one-level submaximal exertion.
Study quality 1 is excellent. Study quality 2 is good. Study quality 3 is poor.
Numbers in parentheses refer to the number of validity coefficients.

Table A.8

Descriptive Statistics for Statistically Significant Two-Way Interactions by Study Characteristics for Oxygen Uptake

Gender × Study Quality

		Gender	
		Males	Females
Quality	1	.730 (39)	.811 (6)
	2	.613 (16)	-.045 (2)

Exercise Type × Protocol

		Type		
		1	2	5
	1	.643 (20)	.624 (8)	.827 (3)
	2	.745 (5)	.148 (4)	.881 (5)
Protocol	3	—	.750 (1)	—
	4	.871 (3)	-.610 (1)	—
	5	.733 (20)	.381 (4)	—

Note. Study quality 1 is excellent. Study quality 2 is good.
Type 1 is bicycle ergometer. Type 2 is treadmill. Type 5 is swimming.
Protocol 1 is progressive continuous. Protocol 2 is progressive intermittent. Protocol 3 is random intermittent. Protocol 4 is one-level maximal exertion. Protocol 5 is one-level submaximal exertion.
Numbers in parentheses refer to the number of validity coefficients.

Table A.9

Mean Validity Coefficients by Criterion Variable and by Study (Overall)

Study	Heart Rate (k[a])	Blood Lactate (k)	Oxygen Uptake (k)	Overall (k)
Allen & Pandolf (1977)	—[b]	.64 (1)	—	.64 (1)
Anshel et al. (1991)	.48 (9)	—	—	.48 (9)
Arstila et al. (1994)	.86 (3)	—	—	.86 (3)
Bacharach (1984)	.65 (2)	—	—	.65 (2)
Bar-Or et al. (1972)	.79 (4)	—	—	.79 (4)
Bloem et al. (1991)	.35 (1)	.61 (1)	.56 (4)	.53 (6)
Borg (1973)	.59 (6)	—	—	.59 (6)
Borg et al. (1977)	.82 (1)	—	—	.82 (1)
Borg et al. (1985)	.50 (8)	.39 (8)	—	.45 (16)
Borg et al. (1972)	.27 (4)	—	—	.27 (4)
Borg, Van den Burg et al. (1987)	1.00 (3)	—	—	1.00 (3)
Brown et al. (1996b)	.81 (1)	—	.83 (3)	.82 (4)
Butts (1982)	.53 (6)	—	-.05 (2)	.38 (8)
Clapp & Little (1994)	.44 (2)	—	—	.44 (2)
Costa & Gaffuri (1977)	.84 (1)	—	—	.84 (1)
Demello et al. (1987)	—	.74 (4)	.69 (4)	.72 (8)
Duncan et al. (1996)	.98 (2)	—	—	.98 (2)
Edwards et al. (1972)	.87 (2)	.70 (2)	.67 (6)	.72 (10)
Gamberale (1972)	.57 (4)	—	—	.57 (4)
Goslin & Rorke (1986)	.47 (1)	—	.75 (1)	.61 (2)
Hardy & Rejeski (1989)	.85 (1)	—	.77 (3)	.79 (4)
Hassmen (1990)	1.00 (8)	—	—	1.00 (8)
Higgs & Robertson (1981)	.67 (4)	—	—	.67 (4)
Jackson et al. (1981)	.16 (1)	—	—	.16 (1)
Kamon et al. (1974)	.48 (1)	—	.59 (3)	.56 (4)
Kay & Shepard (1969)	—	.15 (1)	—	.15 (1)
Kilbom (1971)	.86 (6)	.74 (6)	—	.80 (12)
Kolkhorst et al. (1996)	.61 (2)	—	.15 (4)	.30 (6)
Kurokawa & Ueda (1992)	.74 (8)	—	.92 (2)	.78 (10)
Lamb (1995)	.57 (1)	—	—	.57 (1)
Lamb (1996)	.61 (3)	—	—	.61 (3)
Ljunggren et al. (1987)	—	.48 (2)	—	.48 (2)
Lollgen et al. (1977)	.63 (1)	—	—	.63 (1)
Lollgen et al. (1980)	—	—	.79 (1)	.79 (1)
Miller et al. (1985)	.34 (6)	—	—	.24 (6)
Morgan & Pollack (1977)	.43 (1)	.61 (1)	.52 (1)	.52 (3)
Noble et al. (1973)	.78 (6)	—	.74 (18)	.75 (24)
O'Neill et al. (1992)	.65 (3)	—	—	.65 (3)
Pandolf et al. (1972)	—	—	.69 (2)	.69 (2)

(table continues)

Study	Heart Rate (k[a])	Blood Lactate (k)	Oxygen Uptake (k)	Overall (k)
Pedersen & Welch (1977)	.81 (1)	.80 (1)	.80 (2)	.80 (4)
Robertson et al. (1979)	—	.46 (3)	—	.46 (3)
Robertson et al. (1986)	—	—	.87 (3)	.87 (3)
Rudolph & McAuley (1996)	—	—	-.61 (1)	-.61 (1)
Sargeant & Davies (1977)	.87 (1)	—	.81 (3)	.82 (4)
Sargeant & Davies (1973)	.84 (1)	—	.88 (3)	.87 (4)
Skinner et al. (1969)	.83 (1)	—	—	.83 (1)
Skrinar et al. (1993)	—	.68 (1)	—	.68 (1)
Smith (1994)	.03 (1)	—	—	.03 (1)
Steed et al. (1994)	—	.90 (1)	—	.90 (1)
Stevenson et al. (1982)	.87 (1)	—	—	.87 (1)
Toner et al. (1986)	.68 (1)	—	.56 (2)	.60 (3)
Travlos & Marisi (1996)	.52 (15)	—	—	.52 (15)
Turkulin et al. (1977)	.64 (2)	—	—	.64 (2)
Ueda et al. (1993)	.87 (1)	.76 (2)	.83 (2)	.81 (5)
Ueda & Kurokawa (1991)	.88 (1)	—	.91 (1)	.90 (2)
Ulmer et al. (1977)	.88 (1)	—	—	.88 (1)
Williams & Eston (1988)	.74 (1)	—	—	.74 (1)
Wilmore et al. (1985)	.29 (3)	—	.33 (3)	.31 (6)
Winborn et al. (1988)	.31 (16)	—	—	.31 (16)
Zeni et al. (1996)	1.00 (2)	.99 (2)	—	.99 (4)

[a] k is the number of validity coefficients.
[b] A dash indicates that that study did not report a value.

CURRICULUM VITAE

Michael J. Chen

University of Mississippi Medical Center, Department of Psychiatry and Human Behavior, Guyton Room 128, 2500 North State Street, Jackson, MS 39216-4505, (601) 984-5798; mchen@psychiatry.umsmed.edu

Current Position:

Post-Doctoral Research Associate, Department of Psychiatry and Human Behavior, University of Mississippi Medical Center, Jackson, MS.

Education:

Ph.D. Psychology - Research and Evaluation Methodology (1998)
Utah State University, Logan, UT.
Dissertation Title: *Criterion-Related Validity of the Borg Ratings of Perceived Exertion (RPE) Scale: A Meta-Analysis*

M.S. Biochemistry/Molecular Biology (1995)
Utah State University, Logan, UT.
Thesis Title: *Characterization of an Axial Ligand Substitution in Sperm Whale Myoglobin*

M.A. Experimental Psychology (1984)
San Francisco State University, San Francisco, CA.
Thesis Title: *Structure-Function Relationships of Seven Novel Synthetic Dynorphin Analogues*

B.A. Psychology (1982) University of California, Davis, CA.

Primary Scientific Areas Studied:

*Meta-Analysis - Correlational and Experimental.

*Generalizability Theory - mathematical modeling.

*Program Evaluation - survey design and implementation, focus groups, statistical analysis and reporting.

* Protein Purification - Have experience in purifying and characterizing cellular retinoic acid binding protein, horse erythrocyte catalase, sperm whale myoglobin and several myoglobin mutants.

* Hemeprotein Biochemistry - Determined the structure and function relationships of various sperm whale myoglobin and cytochrome *c* peroxidase mutants using a wide variety of physical and biochemical techniques.

* Recombinant Myoglobin - Molecular biology of purified mutant myoglobins and used x-ray absorption and optical spectroscopies to investigate structural and biochemical changes.

* Teratology of Retinoic Acid and its Congeners - Purification of cellular retinoic acid binding protein from fetal hamster tissue and from bovine testes; production of antibodies against nuclear retinoic acid receptors in New Zeland White rabbits.

* Neuroscience - Investigated the effects of various noradrenergic drugs on recovery of motor function following cerebral trauma in rats and cats.
  -Investigated the effects of cortical contusion on cerebral metabolism in rats.
  -Investigated the changes in adrenergic receptor binding activity during kindling in rats.
  -Investigated the molecular biology of imidazoline receptors in depression.

<u>Recent Teaching Experience (listed by course name):</u>

- Advanced Program Evaluation
- General Pharmacology
- Statistics for the Social Sciences
- Child 2-5
- Infancy
- Introduction to Early Childhood Education

<u>Evaluation Experience</u>:

- Assistant Editor, *The American Journal of Evaluation* (Formerly *Evaluation Practice*)
- Served as a reviewer of manuscripts submitted to the following journals:
  The American Journal of Evaluation
  Evaluation
- Evaluation of the *Research and Evaluation Methodology Program*, Dept. of Psychology, Utah State University (1995).
- Evaluation of the *Parents of Children in the Cache County Migrant Summer Program* (1995).

Employment History:

8/98 to present	Post-Doctoral Research Associate, Department of Psychiatry and Human Behavior, University of Mississippi Medical Center, Jackson, MS.
5/96 to 6/98	Assistant Editor, *The American Journal of Evaluation* (Formerly *Evaluation Practice*).
Intermittent	Consultant in statistics and research design and methods.
Intermittent -ongoing	Consultant in Microbiology, Molecular Biology, Biochemistry. Consulted for Sorensen Bioscience, Midvale, UT.
08/91 to 3/95	Graduate Research Assistant, National Center for the Design of Molecular Function, Utah State University, Logan, UT 84322-4630.
08/88 to 08/91	Graduate Research Assistant and Laboratory Manager (for 6 months), Dept. of Animal, Dairy & Veterinary Science, Utah State University, Logan, UT.
08/85 to 07/88	Laboratory Technician/Graduate Research Assistant, Dept. of Psychology & Physiology and Pharmacology, School of Medicine, University of New Mexico, Albuquerque, NM.
08/83 to 08/85	Laboratory Technician, Dept. of Psychology, San Francisco State University, San Francisco, CA.
08/82 to 08/84	Graduate Teaching Assistant in physiology, psychology and statistics, Dept. of Psychology, San Francisco State University, San Francisco, CA.

Community and Administration:

- Volunteer - Helpline/Crisis/Information Referral, Cache Valley, 1988 - 1995.
- Laboratory Management and Training, Utah State Univ., Dept. Animal, Dairy, and Veterinary Science, 1989-1991.
- Grant-Writing - Successful in acquiring funding from National Institutes of Health, 1989.
- Laboratory Management and Training, University of New Mexico, Dept. of Psychology, 1987.
- Laboratory Management and Training, San Francisco State Univ., CA, 1983-1984.
- Teaching Experience in Clinical Psychology, Physiology and Statistics, San Francisco State Univ., CA, 1982-1984.

- Early Childhood Education Instructor (Montessori), Davis, CA, 1982.

Professional Affiliations:

Society for Neuroscience
Inter-Mountain Chapter of the Society for Neuroscience
Society for Experimental Biology and Medicine
American Psychological Society
American Evaluation Association

Honors Conferred:

Graduate and Professional Opportunity Program Fellowship, Southwest Hispanic Institute,
University of New Mexico, Albuquerque, NM, Aug, 1985 - May, 1986.

National Dean's List 1996-1997.

Unpublished Manuscripts

Chen, M. J., & Hurd, J. (1996). Neurobiology of Alzheimer's Disease: A Review.

Manuscripts in Preparation

Chen, M. J., & Fan, X. The use of Generalizability Theory with time as a resource constraint.

Chen, M. J., & Fan, X. A meta-analysis of the effects of parental involvement on academic achievement.

Manuscripts Submitted

Fan, X., & Chen, M. J. Academic achievement of rural students: A multi-year comparison with their peers in suburban and urban schools. Submitted to: *Journal of Research in Rural Education.*

List of Publications

Published Manuscripts in Refereed Journals:

Chen, M.J., & Fan, X. (In press). Relationship between variance components and the mean difference effect size. *Current Psychology*.

Queen, S.A., Chen, M.J., & Feeney, D.M. (1997). *d*-Amphetamine attenuates decreased glucose utilization after unilateral sensorimotor cortex contusion in rats. *Brain Research, 777*, 42-50.

Fan, X., Chen, M.J., & Matsumoto, A.R. (1997). Gender differences in mathematics achievement: Findings from the *National Education Longitudinal Study of 1988*. *The Journal of Experimental Education, 65*, 229-242

Sinclair, R., Hallam, S., Chen, M., Chance, B., & Powers, L. (1996). Active site structure of cytochrome *c* peroxidase and myoglobin mutants: Effects of altered hydrogen bonding to the proximal histidine. *Biochemistry, 35*, 15120-15128.

Kline, A.E., Chen, M.J., Tso-Olivas, D.Y., & Feeney, D.M. (1994). Methylphenidate treatment following ablation-induced hemiplegia in rat: Experience during drug action alters effects on recovery of function. *Pharmacology, Biochemistry and Behavior, 48*, 773-779.

Chen, M.J., Vigil, A., Savage, D.D., & Weiss, G.K. (1990). Transient elevation of amygdala α_2-adrenergic receptor binding sites during the early stages of amygdala kindling. *Epilepsy Research, 5*, 85-91.

Jimenez-Rivera. C.A., Chen, M.J., Vigil, A., Savage, D.D., & Weiss, G.K. (1989). Transient elevation of locus coeruleus α_2-adrenergic receptor binding during the early stages of amygdala kindling. *Brain Research, 485*, 363-370.

Chen, M.J., Coss, R.G., & Goldthwaite. R.O. (1983). Timing of dispersal in juvenile jewel fish is unaffected by available space. *Developmental Psychobiology, 16*, 303-310.

Published Abstracts and Poster Presentations:

Chen, M., & Fan, X. (1997). Mathematical relationship between variance components and the mean difference effect size. *American Psychological Society Abstracts,* 72. Poster presented at the 9th Annual Meeting of the American Psychological Society in Washington D.C.

Chen, M., Sinclair, R., Hallam, S., Sligar, S., & Powers, L. (1993). Characterization of an axial ligand substitution in sperm whale myoglobin. *Protein Science Abstracts, 2*, 68. Poster presented at the 7th Annual Meeting of the Protein Society in San Diego.

Chen, M.J., Howard, W.B., Jurek, A., Sharma, R.P., & Willhite, C.C. (1989). Expression of cellular retinoic acid binding protein in fetal tissues. *The Pharmacologist, 31*, 139. Poster presented at the Annual Meeting of the American Society for Pharmacology and Experimental Therapeutics in Salt Lake City.

Weaver, M.S., Chen, M.J., Westerberg, V.S., & Feeney, D.M. (1988). Locus coeruleus lesions facilitate recovery of locomotor function after sensorimotor cortex contusion in rat. *Society for Neuroscience Abstracts ,14*, Poster presented at the 18th Annual Meeting of the Society for Neuroscience in Toronto.

Kline, A.E., Flores, T.P., Tso-Olivas, D.Y., Chen, M.J., & Feeney, D.M. (1988). Effects of methylphenidate on recovery from ablation-induced hemiplegia. *Society for Neuroscience Abstracts, 14*, 1152. Poster presented at the 18th Annual Meeting of the Society for Neuroscience in Toronto.

Chen, M.J., Rolland, T.M., Queen, S.A., & Feeney, D.M. (1988). Modification of cerebral hypo-metabolism after cortical contusion: Comparison of glucose utilization rates and cytochrome oxidase histochemistry. *Society for Neuroscience Abstracts, 14*, 996. Poster presented at the 18th Annual Meeting of the Society for Neuroscience in Toronto.

Chen, M.J., Vigil, A., Laubert, C., Savage, D.D., & Weiss, G.K. (1987). Alpha-2 adrenergic receptors in the amygdala are transiently elevated during the early stages of kindling. *Society for Neuroscience Abstracts, 13*, 949. Poster presented at the 17th Annual Meeting of the Society for Neuroscience in New Orleans.

Flores, T.P., Tso-Olivas, D.Y., Kline, A.E., Chen, M.J., & Feeney, D.M. (1987). Effects of methylphenidate on motor ability after ablation injury. The Society for the Advancement of Chicanos and Native Americans in Science, p. 16.

Sutton, R.L., Chen, M.J., Hovda, D.A., & Feeney, D.M. (1986). Effects of amphetamine on cerebral oxidative metabolism following brain damage as revealed by quantitative cytochrome oxidase histochemistry. *Society for Neuroscience Abstracts, 12*, 1404. Poster presented at the 16th Annual Meeting of the Society for Neuroscience in Washington D.C.

Chen, M.J., Sutton, R.L., & Feeney, D.M. (1986). Recovery of function following brain injury in rat and cat: Beneficial effects of phenylpropanolamine. *Society for Neuroscience Abstracts, 12,* Poster presented at the 16th Annual Meeting of the Society for Neuroscience in Washington D.C.

Sutton, R.L., Chen, M.J., Becenti, R., Sandoval, R., & Feeney, D.M. (1986). Effects of phenylpropanolamine (PPA) on recovery of function following brain injury. Minority Biomedical Res. Soc.